# AUTOMATED QUANTITATIVE PHENOTYPING AND HIGH-THROUGHPUT SCREENING IN *C. ELEGANS* USING MICROFLUIDICS AND COMPUTER VISION

A Dissertation
Presented to
The Academic Faculty

by

Matthew Muria Crane

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Bioengineering

Georgia Institute of Technology
August 2011

# AUTOMATED QUANTITATIVE PHENOTYPING AND HIGH-THROUGHPUT SCREENING IN *C. ELEGANS* USING MICROFLUIDICS AND COMPUTER VISION

Approved by:


Dr. Hang Lu, Advisor
School of Chemical and Biomolecular Engineering
*Georgia Institute of Technology*

Dr. Michelle LaPlaca
School of Biomedical Engineering
*Georgia Institute of Technology*

Dr. Robert Butera
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Dr. James Rehg
School of Interactive Computing
*Georgia Institute of Technology*

Dr. Oliver Brand
School of Electrical and Computer Engineering
*Georgia Institute of Technology*


Date Approved: March 10, 2011

To my AsianCajun

## ACKNOWLEDGEMENTS

I wish to thank my labmates for making this process interesting, educational and above all, entertaining. I'd like especially to thank Jeffrey Stirman and Kwanghun Chung for teaming up with me (or allowing me to team up with you), and for the years of coffee breaks and discussions. Ed Park for hours of political distraction and fierce debate. My advisor, Hang Lu, for providing me with all of the feedback and support a graduate student could ask for, while letting me explore this research topic with abandon. Thank you for helping me learn how to research, not just do research. My committee for providing me with valuable feedback and advance to improve this work. My family for supporting and encouraging me and helping to keep me from losing myself in the process. My fiancée, Lauren, for all of her support, forbearance and for making every day brighter than it could have ever been without her.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

## SUMMARY

Due to the large extent to which important biological mechanisms are conserved evolutionarily, the study of a simple soil nematode*, C. elegans*, has provided the template for significant advances in biology. Use of this model organism has accelerated in recent years as developments of advanced reagents such as synapse localized fluorescent markers have provided powerful tools to study the complex process of synapse formation and remodeling. Even as much routine biology work, such as sequencing, has become faster and easier, imaging protocols have remained essentially unchanged over the past forty years of research. This, coupled with the ability to visualize small, complex features as a result of new fluorescent reagents, has resulted in genetic screens in C. elegans becoming increasingly labor intensive and slow because microscopy mainly relies on manual mounting of animals and phenotyping is usually visually done by experts. Genetic screens have become the rate limiting factor for much of modern C. elegans research. Furthermore, phenotyping of fluorescent expression has remained a primarily qualitative process which has prevented statistical analysis of subtle features.

To address these issues, a comprehensive system to allow autonomous screening for novel mutants was created. This was done by developing novel microfluidic devices to enable high-throughput screening, systems-level components to allow automated operation, and a computer vision framework for identification and quantitative phenotyping of synaptic patterns. The microfluidic platform allows for imaging and sorting of thousands of animals at high-magnification within hours. The computer vision framework employs a two-stage feature extraction to incorporate local and regional features and allows for synapse identification in near real-time with an extremely low error rate. Using this system thousands of mutagenized animals were screened to indentify numerous novel mutants expressing altered synaptic placement and development. Fully automated screening and analysis of subtle fluorescent phenotypes will allow large scale RNAi and drug screens. Combining microfluidics and computer vision approaches will

have a significant impact on the biological community by removing a significant bottleneck and allowing large-scale screens that would have previously been too labor intensive to attempt.

# CHAPTER 1: INTRODUCTION

This document attempts to convey work in three significantly disparate fields: neurobiology, microfluidics and computer vision, to create a unified system to address a critical bottleneck in model organism neurobiology. Even as many genetic and neuroscience tools have made significant strides in fidelity, ease of use and throughput, methods for performing fluorescent based screens have remained unchanged. Not only has the technology remained the same, but screens already labor intensive have become even slower as they are performed with increasingly subtle fluorescent reporters and based on more subtle patterns. This work addresses this bottleneck by creating a unified system comprised of a microfluidic device, external control components and a computer vision framework. The development of these novel microfluidic and computer vision tools have been applied to neurobiology research in the model organism *C. elegans*. To introduce these fields to scientists in the computer vision, microfluidic or neurobiology fields, this chapter focuses on providing an overview of the three fields as pertaining to this work. Additionally, an overview of recent work and publications in these fields and work performed by previous students in the Lu lab are included to provide context to the work performed in this thesis.

## *C. ELEGANS* AS A MODEL SYSTEM

Research into the small soil dwelling nematode was initially begun by Sydney Brenner in the 1970's. Selected because of its small size, transparency and ease of cultivation, it has gone on to be one of the most popular and widely studied multicellular model organisms. Despite its small size and apparent simplicity, *C. elegans* has ~20,000 genes, almost comparable to the human genome. Additionally, between 60-80% of *C. elegans* genes (depending on the bioinformatics approach used) have human homologs (2-5). Research into *C. elegans* has yielded significant benefits to science, and in the past eight years three Nobel prizes have been awarded for *C. elegans* research. These awards were for the discovery of programmed cell death(6), interference RNA(7), and use of green fluorescent protein(8). Additionally, *C.*

*elegans* was the first multicellular organism to have its genome completely sequenced; this was completed in 1998(9). *C. elegans* provided a bridge between the study of single cell organisms such as *Saccharomyces cerevisiae* and the significantly more complex *Drosophila Melanogaster (10)*.

## STANDARD METHODS: ADVANTAGES AND LIMITATIONS

The worm community has developed a set of standard cultivation protocols to ensure consistency across labs. Worms are grown on PH buffered agar plates. This agar medium is typically called Nematode Growth Medium (NGM). These plates are seeded using specific strains of E. coli, which the worms use as a food source. The two standard strains of E. coli used for nematode culture include OP50 and HB101, and in all of this work worms were grown on NGM plates seeded with OP50. Depending on the culture temperatures, generation time and dwell time in each larval state varies. The work in this dissertation uses culture temperatures of 15°C, 20°C or 25°C. Varying culture temperatures was used to allow multiple days of experiments and screening from a single round of mutagenesis.

There are a number of genetic methods that have been developed and standardized for the worm community. In particular, screening animals in an attempt to find novel mutants is one cornerstone of modern biology. By modifying an organism's genetic composition, biologists are able to discover new phenotypes and learn the role of specific genes. Screens in *C. elegans* can be performed in two different ways: forward or reverse.

A forward screen is the more standard, or conventional screen. In performing a forward screen a few random mutations are introduced into the genomes of many animals, with each animal having different mutations. This is commonly performed through the use of a mutagenic chemical such as ethyl methane sulphonate (EMS) or by using high doses of UV radiation. These mutations typically eliminate the gene functionality, which is why it is called a knockout approach. In *C. elegans*, homozygous progeny are then examined in an attempt to find a relevant and novel phenotype. Once animals of interest are found, they

are sequenced to determine which gene was affected, and thus the role of that gene can be determined. Because only one quarter of the animals screened are homozygous with the mutation, and it is possible the hit some genes more than once, full-genome coverage typically requires screening >100,000 animals.

The recent discovery of interference RNA (RNAi) in *C. elegans* by Fire and Mello(7) created a new type of screening called reverse or knockdown screening. When an animal is exposed to double stranded RNA (dsRNA), the worm stops producing the protein coded by that gene. This effectively knocks down the gene expression while leaving the gene unaffected. In *C. elegans*, this is relatively simple process because the animal merely needs to be soaked in a solution containing the plasmid. Because the plasmids must be designed in advance, however, and individually applied to animal populations, this can be a more labor intensive process. Additionally, many neurons in *C. elegans* have proved relatively resistant to RNAi knockdown which makes neurobiology screening with this method difficult.

A good comprehensive overview of designing good screens for *C. elegans* was published by Jorgensen and Mango(11). In this review paper, genetic screens are divided into three different categories based on the difficulty and degree of labor required: screens from heaven, purgatory and hell. The simplest screens are largely selection screens where animals not possessing the desired phenotype will be killed or repressor screens where the desired phenotype can be easily seen under a dissection microscope. Screens from purgatory include lethal screens as well as multigenerational screens. The most difficult of all, the screens from hell, include any screen that must be performed using a high magnification compound microscope, and completely immobilize the animal such as essentially all screens involving altered synapse expression. As a result of the years spent studying *C. elegans*, many of the simpler, easier screens have already been performed. At the same time that the development of fluorescent reporters has allowed imaging of increasingly subtle features, screening methods have been virtually unchanged over the past 40 years of research. This has resulted in a significant bottleneck that has slowed research even as other methods such as sequencing have become faster and highthroughput. Furthermore, removing this bottleneck

requires the application of several disparate fields to address a single problem, a challenging task. The work in this thesis focuses on creating a unified system to increase the ease and throughput of these challenging screens.

## USING *C. ELEGANS* AS A MODEL FOR NEUROSCIENCE AND SYNAPTOGENESIS

Despite its small size, *C. elegans* has a fully functioning nervous system comprised of 302 neurons. This small nervous system makes it ideal to study as an organism with the goal of understanding the nervous system in its entirety, and the progression from genes to neurons to behaviors. This small nervous system is contrast to approximately 100,000 neurons that fruitflies have and nearly 100 billion neurons present in humans. A significant additional advantage of using *C. elegans* for neuroscience work is the fact that *C. elegans* also has a stereotyped development and neuronal connectivity. Knowledge of the neuronal wiring provides researchers with an unprecedented ability to perturb the network and observe its response.

In 1986 White(12) published a large work which detailed the connectivity between almost all neurons within an N2 hermaphrodite. By analyzing large numbers of electron microscopy images of animal cross-sections White was able to detail the locations of the soma, the number of synapses and dendrites and the connectedness. Utilizing this preliminary map of the neuronal network, researchers began to study the network by eliminating individual nodes. This was accomplished by using tightly focused laser beams to ablate single cells. This has allowed biologists to remove specific nodes of the network and analyze how the network system is able to respond and correct for the absences. Additionally, the development of femtosecond lasers has further limited the amount of damage to surrounding tissues during ablations. This has allowed axon cutting and synaptic ablation, further increasing the number and diversity of experiments that can be conducted.

The discovery of GFP and use as a reporter driven by specific promoters had specific relevance to the study of *C. elegans* neurobiology. Due to *C. elegan*'s transparent tissue, fluorescent proteins are of

particular use as they allow easy visualization and categorization of animals based on the expression pattern of a particular GFP reporter transgene in addition to the behavior or appearance. Using specific promoters upstream from the genetic sequence for GFP, it is possible to visualize when and where specific genes were turned on. Fluorescent based phenotypes have begun to dominate the study *C. elegans* neurobiology, especially in the discovery of the roles of specific genese and mutant discovery. By tagging the fluorescent proteins with a specific protein, they can be trafficked to, and localized at specific regions of the neuron such as the pre- or post-synapse(13). This has allowed the study of where synapses are localized, how they are removed and remodeled, and what extra- and intra-cellular cues guide synapse localization.

## KNOWLEDGE OF EXTERNAL CUES FOR SYNAPSE FORMATION AND MATURATION

Over the past thirty years a large amount of research has been performed focused on understanding synapse function, and more recently, focused on how synapses are formed(14). Synaptogenesis is the process in which two neurons develop a set of release and receptor sites for neurotransmitter release and detection. *C. elegans* proves to be an ideal organism for studying synaptogenesis as it has a highly stereotyped neural connectivity and synapse are formed in repeatable patterns in wild-type animals(15). Additionally, the transparent nature of the animal makes it easy to dissect neural patterns using the integration of fluorescent reporter transgenes(13).

Synapses are characterized by a pre- and post-synapse. The pre-synapse is characterized by small vesicles that contain neurotransmitters. When required, these synaptic vesicles move to, and fuse with the membrane to release the neurotransmitters into the small gap between the pre- and post-synapses. The neurotransmitters diffuse across the gap and are detected by specific ligand-gated channels in the post-synapse. Although the pre-synapses of worms appear structurally similar to other higher organisms, the

post-synapses contain no electron-dense structures – making assignment of post-synapses difficult. Thus, they are typically assigned based on proximity to pre-synapses(14).

In studying synapse formation the creation of a synaptobrevin GFP fusion reporter has proved invaluable. By modifying the transgenes to include not just the promoter and GFP sequences, but also the sequence for an integral membrane protein of synapse vesicles – it became possible to localize fluorescence specifically to synapses(13). This resulted in the ability to screen animals for altered synapse patterns, and investigate the role that specific genes play in synaptogenesis. Since the advent of the original SNB-1::GFP reporter, other fusion transgenes have been developed for visualization of both the pre- and post-synapses(13, 16, 17). These reagents have been successfully used to study large numbers of previously unknown genes. Although these reagents have become widely used, the development of new phenotyping methods has lagged considerable and need to be developed for the characterization of synapse/GFP localization.

There are still significant knowledge gaps regarding the formation of synapses. Synapse formation involves a number of both positive and negative cues; inhibition of synapses where they shouldn't form, and promotion of formation where they should. In particular the model of HSN neurons has been a productive model in uncovering the role of immunoglobin superfamily proteins SYG-1 and SYG-2 and their role in directing synapse formation(18). Additional work was done on the DA9 motor neuron system to uncover some of the negative cues that inhibit synapse formation (19-21).

Although the advent of reagents that allow biologists to accurately pinpoint synapse locations has been hugely beneficial, these markers have limitations in that they are dim and prone to quick photobleaching. Most conventional methods require extend imaging times that can distort, or suppress these signals. Additionally, synapse patterns can be complex, and there is the need for quantitative methods to phenotype these fluorescent patterns as it is currently difficult for researchers to identify subtle phenotypes. Quantitative phenotyping methods that have been applied to synapse expression only

quantify three or four features, and because they are based on user drawn line scans will vary from user to user (22-25). Furthermore, a lack of quantitative phenotyping makes it difficult to compare strains discovered in different labs, and makes it impossible to automate the discovery process.

## MOTOR NEURON DA9 AS A SYNAPTOGENESIS MODEL

Although some work has been done that focuses on the inhibition of synapse formation, the majority of research has been directed at finding positive cues that promote synapse creation. Like most biological systems, however, synaptogenesis is composed of positive and negative cues to ensure precise control and feedback. The cholinergic motor neuron DA9 located in the tail of the animal has recently been used to study synaptogenesis, specifically the role that Wnt signaling plays. A schematic showing the position of important features is shown in Figure 1. The soma of each DA neuron resides near the ventral midline and extends an axon posteriorly and around the dorsal side to innervate the muscle, and also receive inhibitory signals from the complimentary VD motor neurons. A dendrite is sent anteriorly to receive command signals from the sensory and command neurons. Synapses along the dorsal axon are precisely and consistently located, and there is an asynaptic region along the beginning of the axon and in the commissure that crosses the body.

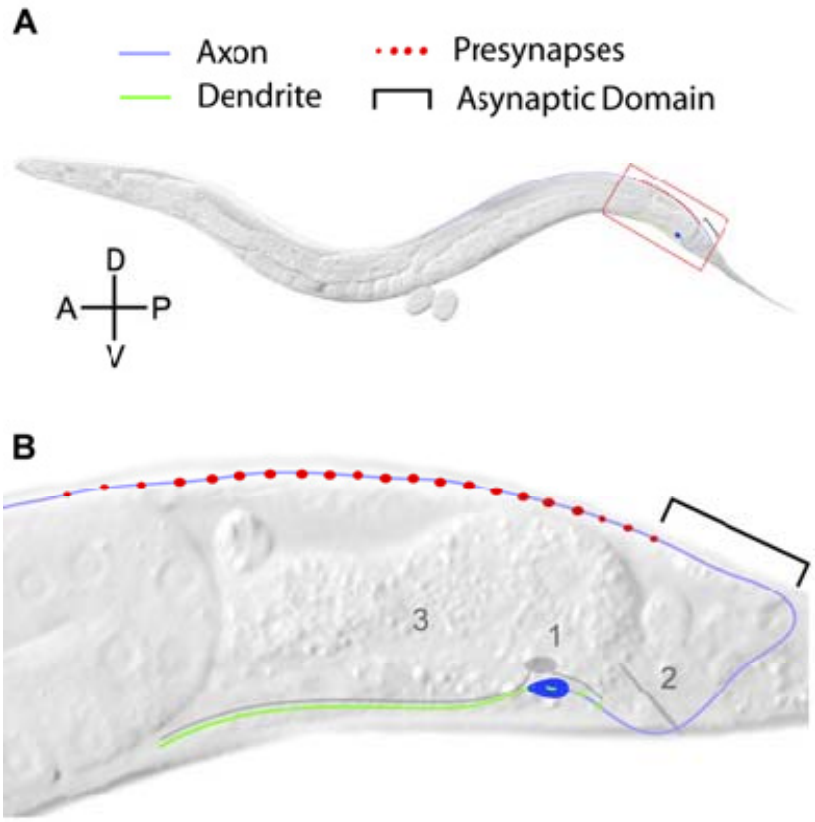



**Figure 1: Image of DA9 axon and synapse localization from Klassen et al.(1)**

Using fluorescent reporters driven by different promoters and fused with different pre- or post-synaptic proteins, researchers were able to study the precise locations of the synapses in different mutants. Rather than doing a forward or reverse screen to search for novel genes, they pursued a candidate approach. Using this method animals containing mutations in genes hypothesized to be involved in the synaptogenesis pathway were selected and phenotyped. Because the phenotyping process required imaging multiple animals under high-magnification (63X or 100X), it would have been prohibitively difficult and time consuming to screen for novel mutants in a forward or reverse screen. Additionally, as the reporters are only localized at the synapses, they photobleach extremely rapidly, and require long exposure times. Using this approach, they identified the *wnt* signaling pathway composed of the *wnt*

molecule *lin-44* and receptor *lin-17* as preventing the formation of synapses in the asynaptic domain. Although the candidate approach proved successful, 10-20 additional genes are likely involved in this pathway and will remain a mystery unless discovered using a screen. The difficulty of performing such a screen, however, makes it prohibitive unless novel methods are created, and most screens performed on these model systems are far from saturated, having been performed on only a small percentage of the genome

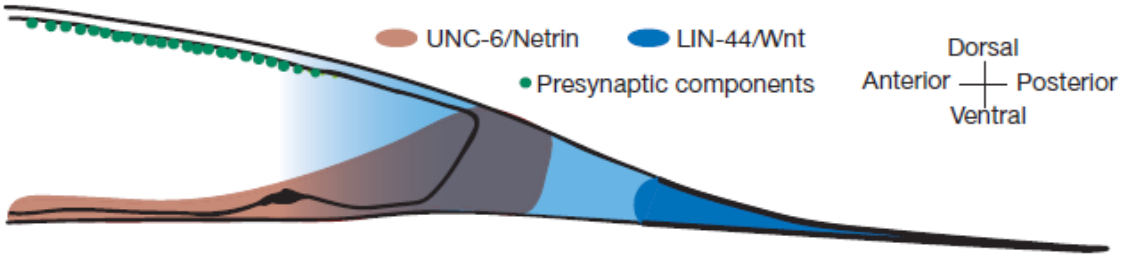



**Figure 2: Several of the inhibitory cues secreted by cells nearby DA9 to prevent the formation of synapses in specific regions(21).**

## FLUORESCENT PHENOTPYING METHODS

Phenotyping is a critical element to the study of model organisms, as it allows researchers to identify specific features that are different between the mutant of interest, and a wild-type animal or population. Once a mutant has an observed phenotype it is possible to relate the specific disrupted gene to a behavior or appearance, and thus hypothesize as to the specific gene function. In neurobiology, the majority of phenotyping is currently done using fluorescent proteins.

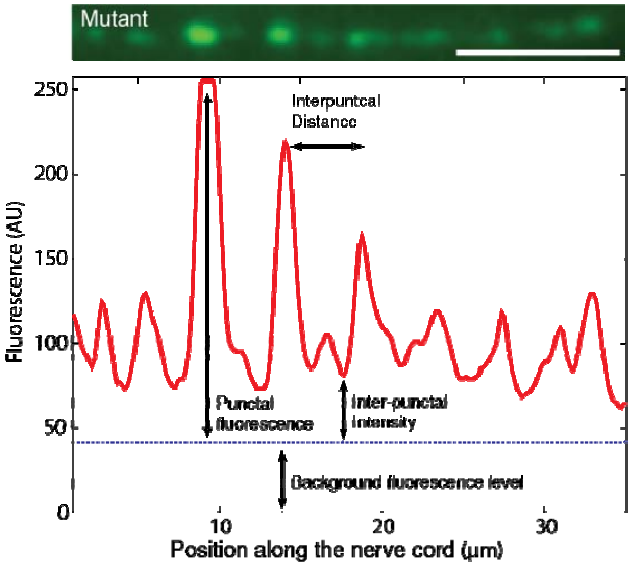



**Figure 3: Line plot along the nerve cord showing current phenotyping methods. Scale bar 10 µm.**

In *C. elegans* work fluorescent phenotyping has by and large been qualitative, but it is increasingly focusing on quantitative descriptions that are necessary for more subtle phenotypes. As more and more research focuses on fluorescent phenotypes, the genes that create obvious distortions in the fluorescent reporters, and it becomes increasingly difficult to identify new players in specific pathways. This problem has spurred the development of increasingly quantitative phenotyping methods, especially in the study of synaptogenesis. Genes involved in these pathways may cause synapses to be created at different locations, reduce or increase the intensity of fluorescence, or result in many other subtle modifications. If the animals cannot be phenotyped, and shown to differ from the wild-type population, than genes within the pathway cannot be discovered. Several excellent example systematic approaches to phenotyping synapse expression have already been demonstrated(25, 26). These methods, however, have to performed by hand in a program such as Image J, requiring at least several minutes per animal – making screens for subtle changes almost unthinkable. Additionally, each additional feature that is added significantly increases the complexity as it moves the phenotyping to a higher dimensional space. This is a

complicating factor as ultimately the decision making is performed by a graduate student or postdoc, and subject to human biases and error.

Line scans along the axon or dendrite in question allowed the researchers to quantify the synapse intensity, the level of fluorescence present between synapses, and other interesting features. The image in Figure 3 is from this classic paper(26), and shows how the various components were computed. While this is currently the standard method used to analyze synapses, only three or four features are analyzed, and the process involves considerable user input leading to significant variation and biases(22-25). This idea was further extended to quantifying the number of synapses present in specific regions, and looking for animals that differed in the number or placement of synapses(20) as can be seen in Figure 4. In this case software used manually to create a line scan, and then the number of synapses in each region was counted by the user, which introduces user bias. This makes it slow and labor intensive and it would be impossible to screen for animals if the phenotyping step for each animal took many minutes. Additionally, a host of features that could be analyzed relating to the size, shape, and placement of synapse are ignored in the current method. Because of these concerns, there is significant interest in utilizing machine learning and computer vision methods to allow for rapid and high-content phenotyping in biological systems. By implementing a computational framework capable of doing this, it would become possible to screen for combinations of features that would be too subtle or challenging to perform visually. Utilizing machine learning and computer vision approaches to identify synapses, would allow for analysis and phenotyping with speed and accuracy to improve throughput and allow automated screening.

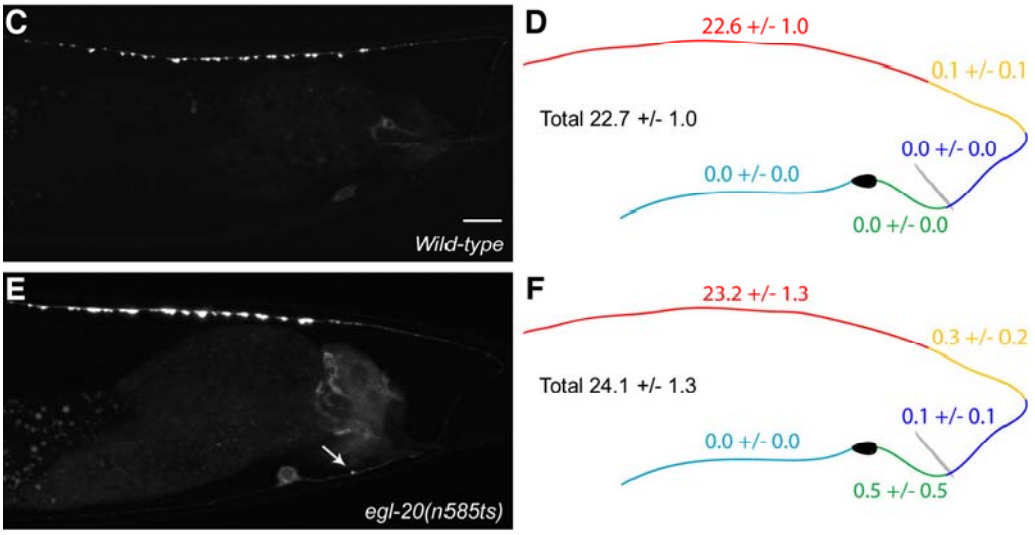



**Figure 4: Phenotyping synapse expression by counting the number of puncta within specific regions(20)**

## CHALLENGES FACING *C. ELEGANS* NEUROBIOLOGY

The advent of multiple novel reagents that have enabled imaging of individual synapses, and even the trafficking of individual synaptic vesicles, has created new challenges for C. elegans neuroscience community. These reporters opened up new avenues of research by making it possible to visualize sub-cellular features, but have simultaneously challenged researchers that rely on microscopy techniques based around manual mounting of animals and phenotyping that is visually done by experts. Rather than screens performed using stereomicroscope, where the throughput could be hundreds of animals an hour, these screens using small, easily photobleachable reporters are limited to throughputs of around a hundred animals an hour(11). This makes it especially challenging to saturate screens and to identify all of the genes involved in specific pathways. An expression of the difficulty of performing forward screens is that researchers sometimes uses a candidate approach(20). In this, the reporter of interest is crossed into specific genetic backgrounds that are suspected to affect the pathway of interest which avoids the difficulty of performing forward screens, but is incapable of identifying new genes.

Even as much routine biology work, such as sequencing, has become faster and easier, imaging protocols have remained essentially unchanged over the past forty years of research. The reliance of visual phenotyping to identify pattern changes has worked well, but is limited by the human ability to quantify and detect pattern changes. Computer vision and machine learning methods have recently been applied to a number of biological problems to allow for high content analysis of cellular morphology and phenotypes to extract large amounts of data, and then analyze it for previously unknown relationships (27-32). Converting the current qualitative phenotyping of fluorescent reporters in *C. elegans* to a quantitative high-content space could allow researchers to identify much more complex interactions. This could allow genes that are affect specific pathways, but in subtle ways, to be identified and characterized.

In summary, the challenges are twofold:

1) Fully saturating screens to identify all of these genes involved in specific pathways
2) Developing quantitative phenotyping methods capable of identifying pattern changes that in mutants that would be too subtle to be identified visually

## MICROFLUIDICS

Microfluidics as a field is a relatively recent development, spurred by several key technological developments and the benefits offered by microfluidics. The length scale of microfluidic devices are engineered on the same scale as many biological systems such as cells or small organisms. This allows researchers to create conditions that would be impossible with larger, conventional systems. Also, flow in such systems is typically laminar and thus easy to control and predict. Because of the small length scales, mass and energy transfer occur much faster(33, 34) than in larger, macroscale systems. These features allow devices to manipulate smaller volumes of expensive reagents with greater precision resulting in lower cost, and potentially more significant results. Originally microfluidic devices were created using conventional hard silicon MEMS. While beneficial, these devices were expensive both in material and

labor costs. One particular technological development was the development of PDMS as a material for soft lithography(35-37) and the creation of a robust, easy to manufacture valve that can be integrated on-chip(38). This has allowed researchers to create complex systems capable of being computer controlled and automated(39). PDMS allowed rapid development and prototyping.

## MICROFLUIDICS AS A TOOL FOR BIOLOGY

As a general tool for biology, microfluidics offers numerous benefits to the biology community. As previously mentioned, the length scale of microfluidic manipulation allows previously unheard of control of chemical and physical properties on the same scale as individual cells. Cells experience complex gradients when in-vivo, and the sensing of specific stimuli and the signal transduction networks is of interest to everyone from neurobiologists to cancer researchers. Prior to the development of microfluidics, it was impossible to generate precisely controlled gradients, or to temporally modulate gradients. Additionally, microfluidics allows shorter mixing times and thus a greater uniformity in individual experiments, and much shorter time-points from stimulation than previously possible.

## BENEFITS OF MICROFLUIDICS FOR MANIPULATION AND AUTOMATION

Many of the advantages conveyed by microfluidics have to do with the length scale, and how fluids behave on that scale. One of the central elements to microfluidics is how fluid dynamics and physics changes significantly during the transition from macro- to micro-scale. The changes in how fluids perform can be extremely non-intuitive, but when used properly can provide advantages not possible with larger scale systems. Dimensionless numbers are often used to convey how the flow is operating and where the specific flow sits in the parameter space.

The most commonly discussed dimensionless number is the Reynolds (Re) number. This conveys the relationship between inertial and viscous forces, and is calculated using the following formula: Re = $\frac{\rho V L}{\mu}$.

Conventional, macroscopic fluids have high Reynolds numbers, whereas virtually all microfluidic systems have low Reynolds numbers and operate in the laminar flow regime. This means that inertial forces are virtually irrelevant, and result in extremely predictable and deterministic flow.

The Peclet number describes mixing and conveys the relationship between the convection and diffusion, and can be calculated using the following formula: Pe = $\frac{VL}{D}$. In macroscale systems the turbulent flow that is typically present creates many random eddies that fold and mix the fluid. In these cases the convective mixing dominates and allows fluids to mix extremely. As a result of the low Reynolds number at which microfluidic systems operate, however, convection is extremely limited and this means that diffusion dominates. In these cases, mixing between two fluids can take an extremely long amount of time depending on the coefficient of diffusion of the particles in question. This can prove extremely advantages, for example to flow a step function of two separate chemicals over a cell, where no mixing occurs. Or it can be disadvantageous, when mixing is desired. Significant amounts of work in the microfluidics field have focused on the need to create robust and reliable mixers.

Similar to the development of microelectronics, research into microfluidics has been based around the idea that microfluidic devices can provide a significant advantage to laboratories by allowing automation and the gathering of higher quality data. The development of large scale valve integration, and the incredibly small reagent amounts required in a microfluidic system, has made it possible to perform hundreds of experiments on a single device(40, 41). On a single microfluidic chip, hundreds or thousands of chambers can be created and designed so that they each expose the cells contained in the chamber to slightly different concentrations of multiple chemicals. This promises not only to save significant amounts of money due to reduced reagent use, but also to allow much larger scale biological experiments to be conducted within a reasonable timeframe.

The physical properties that can be controlled by microfluidic systems offer the opportunity to create experiments where specific parameters are controlled with far greater accuracy and precision than would be possible using macroscale instrumentation. For example, investigations into specific cell signaling pathways may require a cell line to be exposed to a specific chemical for a period of time, then fixed, lysed, and specific protein levels analyzed(42). Conventionally this would be done by pipetting a specific volume of the chemical into tube where the turbulent flow would cause mixing. Because each cell is exposed to the chemical at slightly different times depending on where it is in the solution, this results in a range of responses when attempting to measure cellular responses to only a few seconds of exposure. Using a microfluidic system, and the deterministic flow, it is possible to design a system that ensures all cells are exposed to the stimulus at the same time. Thus, not only can microfluidics spur laboratory automation and higher throughput, but can be used to design experiments that were previously not feasible.

## GENERAL MICROFLUIDIC METHODS

The current microfluidic methods were primarily developed by Whitesides(35) and Quake(38), and rely on the premise of micromolding. Creating devices through direct machining or milling can be useful, but by and large these methods have limited resolution and fidelity, which restricts their use to devices requiring channels hundreds of microns large. Conventional hard-silicon MEMS technology provides sufficient resolution, but is expensive and extremely time consuming. Rather than attempting to micro-mill or etch each device using conventional MEMS technology, the most current microfluidics use a method called soft-lithography(36). This has a number of advantages over more conventional MEMS or semiconductor processing and fabrication methods: it is less expensive, it allows the fabrication of simple valves and turnaround times are significantly faster. Soft-lithography as it is typically called, involves the creation of a master which contains the negative of the final device. The master is typically created using a blank silicon wafer with 3-dimensional structures created on top of it made out of photoresist by micropatterning

techniques. The device is then molded by pouring an uncured polymer (typically PDMS) on top of the master and curing it. Once the curing process is completed, the devices are removed, and the master can be reused. Individual PDMS devices are then placed in a plasma chamber to activate the surface which allows them to irreversibly bond to glass slides. This creates a strong seal and allows high pressures to be applied during operation. Should multiple layers be required, they can be fabricated in a similar fashion, and bonded by partially curing the individual layers, connecting them and then completing the curing process. A mold of this master is then created using a soft polymer (PDMS) which is poured over the master and cured. This polymer mold is finally removed and used. One significant advantage of this method is that the molding process is simple and doesn't harm the master, so a single master can be used to make thousands of devices.

Early microfluidics had been hampered by the ability to simply and robustly create valves on-chip that could manipulate and control fluid flow. Although passive devices designed around basic physics principles are capable of solving many problems and can be simpler and more elegant solutions than the active alternatives, active on-chip valves are a necessary component for long-term microfluidic objective of creating devices capable of automating many laboratory processes. The initial microfluidic valve attempts used conventional MEMS substrates such as Si, SiC or SiN(43). These valves were both extremely costly in time and materials to produce, and hybrid polymer-Si devices weren't that much simpler to use(44, 45). The advent of soft-lithography and micro-molding, however, created a medium that allowed a simple valving mechanism based on the deformation of a thin membrane into the channel, analogous to a pinch valve(38). That allowed large numbers of valves to be integrated simply and easily into a small region on a microfluidic device(39). These valves are created by crossing a fluid line with a control line in a separate layer, with a thin membrane separating the two. When the control line is pressurized, it deforms into the flow layer eliminating flow. By changing the pressure applied in the control layer, the valve can be

closed completely, or the cross-sectional area of the flow layer can be reduced. The various operation modes of these valves can be seen in Figure 5.

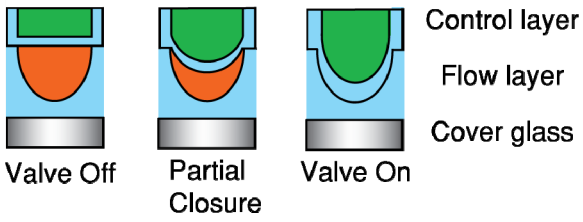



**Figure 5: Schematic showing the operation of a full-closure valve. Using a curved flow layer created by reflowing the photoresist allows the membrane from the control layer to deform and completely close the flow layer.**

## STATE OF THE ART MICROFLUIDICS FOR MULTICELLULAR MODEL ORGANISMS

Using microfluidics to perform experiments on multicellar model organisms is a relatively recent idea, and owes significantly to the work performed within Dr. Cori Bargmann's lab at Rockefeller(46-48). To push the frontier of using microsystems for multicellular organisms to truly benefit genetics and therapeutic screens, a few challenges need to be overcome. First, because of the large number requirement for these types of studies, the micro systems have to be robust over long periods of time; the practicality of running a system thousands to millions of times is very different from running a one-time assay. Second, because some of these systems need to recover live samples (such as in a genetic screen), not just the information gathered from assaying a biological system, the chips need to be gentle and be able to sustain the organisms' physiology (and growth in some cases). Third, many genetic, cell biology studies, and drug screens are morphometric studies, largely using fluorescent markers or reporters. For reasons stated earlier, phenotypical analyses also benefit from quantitation and automation. Therefore the integrated system, the control scheme, the analyses, and the hardware system all have to be compatible and work well with the existing optical microscopy setups. Recent advances in these areas are presented below.

One method of environmental control is manipulation of the geometrical environment and/or the mechanical properties of the materials. Controlling the physical properties surrounding the organism, through the modification of the geometry and material stiffness, allows researchers to investigate the effect of the local environment on behavior and locomotion. For example, *C. elegans* moves by crawling in a sinusoidal manner and generating thrust for forward or backwards motion. This process relies on both the material and geometric properties surrounding the organism. Two groups have used similar pillared structured arrays to control, modulate, and observe the crawling pattern of *C. elegans* as affected by its environment(49, 50). By altering the pillar size and pillar to pillar distance was able to measure a difference in crawling velocities. These microstructured devices are easy to make using PDMS or agar replica molding from a master and offer a simple and inexpensive method to investigate how the shape and material properties of an organism's microenvironment affect its locomotion. These devices are now in the position to allow for complex analysis of behavior genetics and mechanosensory or chemosensory biology.

## MICROFLUIDICS FOR CHEMICAL DELIVERY AND BEHAVIOR

In addition to mechanical environment, multicellular organisms have to integrate complex chemical cues in order to survive, and thus the ability to directly control and selective expose organisms to chemical cues is of great importance. For example, chemotaxis assays are commonly performed in macro-scale environments (e.g. using an agar plate with odors or tastants spotted in one part of the plate), but the standard techniques offer researchers limited control of concentration and gradient quality, and result in a large degree of uncertainty when behavior results are analyzed. By capitalizing on laminar flow, small diffusion time scales, and spatial confinement, researchers have found microfluidic technologies key to improving standard assays through more controlled gradients and faster temporal control. In an early work, Gray *et al*. probed the ability of *C. elegans* to sense molecular oxygen(51). The researchers created a stable and repeatable gradient of dissolved oxygen by flowing nitrogen and air (21% oxygen) through

opposite ends of a PDMS device as “sink” and “source” for the gradient. Creating an oxygen gradient within a confined area allowed for probing *C. elegans* response to a much more precisely controlled gradient and in a much higher throughput manner than previously possible.

Another example of handling gasses and establishing odor gradients for longer-term experiments is the maze learning assay(47). By combining spatial or geometric design restrictions with the constraints these designs placed on the diffusion of chemical signals has allowed researches to probe *C. elegans* olfactory chemotactic response to pathogenic bacteria. This device consisted of a central chamber connected to eight channels leading to open chambers where different strains of pathogenic bacteria were spotted. The worms were then placed in the central chamber where they experienced multiple chemical cues from the various bacteria used. By observing the choices of the worms, olfactory learning of trained and naive, wild-type or mutant *C. elegans* can be measured; this further allows the dissection of the genetic pathways and neural circuitry of olfactory learning. Compared to an open-plate conventional chemotactic assay, this olfactory maze allows for directional cues (i.e. not complete mixing of all the odors), which is critical in allowing for the “multiple-choices” instead of the simple traditional two-choice assays.

In addition to creating gas gradients, microfluidic technologies also lend themselves to fast and repeatable switching between dissolved gasses of constant concentrations. PDMS has a relatively high permeability to oxygen as well as other gasses. Using a simple two-layer PDMS device Zimmer *et al.*(52) imaged neuron activities (calcium transients using calcium-sensitive fluorescent proteins) in response to temporal step gradients of oxygen(52). In switching of oxygen concentrations, it was found that the dissolved oxygen equilibrated in 5-10 sec after switching and this allowed for both steps-up and -down in concentrations and the recording of the neuronal response in a highly repeatable manner. PDMS gaseous permeability has also been utilized by other researchers for immobilization of *C. elegans* (which will be discussed later in this review).

The exploitation of microfluidic laminar flow as applied to multi-cellular model organisms can be best seen in the work by Chronis *et al*.(46). In this work the researchers wished to deliver a chemical stimulus across the nose of *C. elegans* (where many sensory neurons have exposed ciliated processes) and record neuronal activities measured by calcium transients. In order to investigate the response rate and magnitude of chemosensory neurons to a variety of chemicals, the stimuli needed to be delivered with both precise temporal control. In a microfluidic device, the worm was loaded into a channel with a smaller cross-section than that of the animal to restrict movement; it was held in place by positive pressure and temporally stimulated by the odors of interest. The temporal control of the simulation was achieved by opening or closing the flow to the control side channels, which also maintains the overall volumetric flow rate and thus the pressure. To ensure that no mixing occurred between the buffer and stimulus flow streams, a relatively high flow rate of buffer and stimulus was used (i.e. high Peclet number for mass diffusion).

Fakhoury *et al.* was able to use a simple microfluidic device for the analysis of drug affects on drosophila embryo development(53). A Y-channel design was used to introduce one or two drugs into the main channel. In this main channel, a number of embryos were immobilized using interfacial tension created by oil, water/alcohol and SAM-modified surfaces, and a single or combination of drugs was flowed over the embryos. The constant flow of drug ensured a consistent concentration, and the small channel dimensions allowed for minimal reagent consumption. Again, a set of relatively simple to fabricate and use devices has allowed researchers to probe the response of an organism on time and spatial scales not possible using macro methods while using smaller amount of reagents.

Besides controlling mass transport, laminar flow can also be used for controlling heat transfer and thus temperature distributions in microfluidic devices. Luchetta *et al.* and others used a simple Y-shaped PDMS microfluidic device in which one inlet contained a warm buffer solution and the other a cold solution to control the developmental rate in different parts of live *Drosophila* embryos(53-55). Because

of the laminar flow and relatively large flow rates (resulting in high Peclet numbers for heat transfer, analogous the mass diffusion scenario discussed earlier), little thermal diffusion occurred across the two streams and a sharp temperature step was created across the *Drosophila* embryo. By visualizing the number of nuclei on the two halves of the embryo that were exposed to the two different temperatures, the researchers were able to discern developmental differences in the same embryo due to temperature effects and begin to understand how embryos control such important processes. This work was later expanded with the parallelization of many fly embryos in a temperature step gradient to further increase the throughput of such assays(56).

Thermal control in microfluidics has also been used to immobilize the nematode *C. elegans*(57, 58). *C. elegans* has very small thermal mass: an adult animal of ~$10^{-5}$ J/K and L1 larva of ~$10^{-7}$ J/K; in other words, to raise the temperature by 10K, 0.1 mJ of heat is needed for an adult animal and 1 µJ for an L1 animal. Due to this extremely small thermal mass, *C. elegans* can be cooled or warmed very rapidly (practically instantaneously) by controlling the surrounding temperature. Chung *et al.* incorporated a cooling channel into the devices, relying on thermal diffusion through a thin PDMS membrane (also of small thermal mass) between the temperature controlling channel and the worm channel to achieve rapid immobilization. The animals were instantaneously immobilized as soon as they entered the cooling region at ~4 °C, and instantaneously mobilized as soon as they left the region to warm back to room temperature.

## MICROFLUIDICS FOR IMMOBILIZATION

When it comes to handling small model organisms, the most unique advantage conferred by microfluidic technology is integration of functional components. On-chip functional components could enable sophisticated level of control that is otherwise impossible to achieve using macro-scale methods. For example, in order to image *C. elegans*, one needs to immobilize samples with anaesthetics, manually mount the samples on a slide glass, locate each individual, and then image them. For phenotypical

screening and laser ablation purposes, additional processes are required, and these include locating target neurons, laser firing, and rescuing the worms by sliding the coverslip off and picking the worms using a "worm pick", which is often a platinum wire. This painstaking manual handling not only significantly limits the experimental throughput, but also increases noise due to variation in sample-to-sample handling.

To address this limitation, three groups independently developed multilayer PDMS devices that can rapidly route, load, and immobilize live *C. elegans* for high-resolution imaging and laser ablation (1, 57-62). These microdevices are similar in utilizing intricate sequences of on-chip PDMS valves(38, 63) to control a buffer suspension of worms, while differing in key mechanisms. Each group explored a unique approach for single worm loading. Zeng and Rohde *et al.* developed two-step suction mechanism whereby a single worm is first captured by a single suction channel while the remaining un-trapped worms are flushed away(59, 61). The captured worm is then transferred from the single suction channel to an array of suction channels on the opposing wall. This is done by manually actuating on-chip valves. Guo *et al.* utilized a set of side manipulation channels located at each end of the imaging/surgery region to manually position a worm in the region(60). Both of these multi-step processes require additional image acquisition, analysis, and valve actuation to coordinate the activities on-chip. An alternative to these designs was designed Chung *et al.*, which uses self-regulated loading mechanism and passive positioning mechanism using partially closable on-chip valves(57, 58). This particular design completely eliminates the need to have (manual or image-based) feedback control of the sample loading to ensure one and only one worm is loaded at a time, and thus ensures robust and consistent operation of the chip. This device will be discussed at length in the following section as Kwanghun Chung and I worked on developing the system together.

Once a worm is loaded, it needs to be completely immobilized because, unlike embryos, *C. elegans* has high motility that can cause blurring of image or incorrect laser ablation. Several groups working in

parallel to this thesis have developed different strategies to achieve immobilization. While Hulme *et al.*(64) and Allen *et al.* (65) uses simple geometric constraint, Guo *et al.*(60) and Zeng *et al.*(61) used the elastomeric properties of PDMS and positive pressure to physically restrain the animal. The latter groups integrated a thin PDMS membrane, similar to an on-chip valve(38), over the channel where the worm is loaded. By pressurizing the membrane, and deforming it around the animal they were able to mechanically restrict the worm's movement reversibly. Both groups demonstrated immobilization of worm's body movement and successfully performed femtosecond-laser microsurgery.

Two other approaches for immobilization are explored by Chung *et al.* and Chokshi *et al.*(57, 58, 62). The two technologies take advantage of the rapid heat and mass transfer in microsystems. As the size of microstructure is reduced, the surface area-to-volume ratio increases, which allows the microsystem to reach steady state rapidly. As reviewed earlier, Chung *et al.* used an integrated temperature control channel to locally cool the loaded worm for immobilization. The small thermal mass of *C. elegans* and huge surface area to volume ratio of the microchannel results in nearly instantaneous immobilization of the body as well as stopping the pharyngeal pumping, which is critical in cell laser ablation. Analogously, Chokshi *et al.* use rapid mass diffusion of gas through PDMS membranes to create a high $CO_2$ micro-environment(62). Although more time for immobilization than the other methods is required (within 1~2 min), this technique proved sufficient and is an alternative for long-term immobilization (1-2 hours).

## LIMITATIONS OF CURRENT SYSTEMS FOR FULLY AUTOMATED SCREENING

These systems presented all confer individual advantages and disadvantages that are largely a result of the inherent tradeoff between the increased abilities of more complex systems, and the increased failure rates/modes that are accompanied by the increases in complexity. The simplest systems, such as those designed by Hulme et al, are the most transferable to biological collaborators as they require few external components, and use a single layer device. This simplicity also reduces the failure modes, for example, a

single layer device cannot delaminate between the PDMS layers, nor can off- or on-chip valves fail to actuate. The downside to systems like this, however is that they incapable of sorting, or most automation. Performing automated screening requires the ability to introduce error-handling routines that control for varations as the system operates. This, of course requires both on-chip valves, and extensive off-chip components and error-handling routines. None of these systems are (at this time) capable of performing such a task.

## PREVIOUS MICROFLUIDICS WORK WITHIN THE LU LAB

Within the Lu lab, one graduate student (Kwanghun Chung) worked on the problem of developing microfluidic devices capable of performing high speed imaging and sorting of *C. elegans* in a consistent manner. In the following section I will highlight some of the key contributions he made to the field, as well as some of the weaknesses with the system that required further development. The paper based on the work was co-first authored by Kwanghun Chung and myself. Dr Chung created and designed the microfluidic devices used in the work, while I created the external components and computer software.

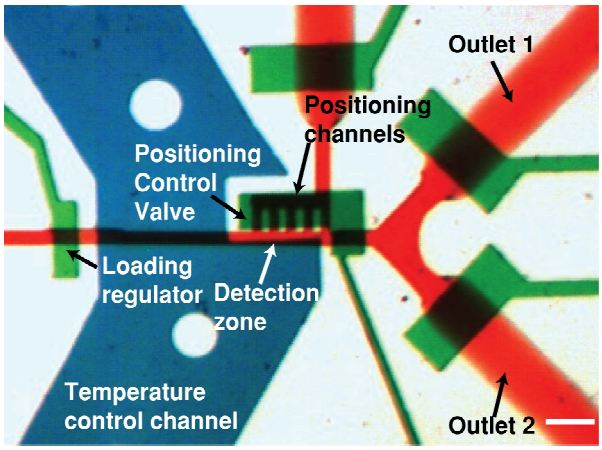



**Figure 6: Optical micrograph of the microchip's active region. The channels were filled with dye to show specific features: blue, temperature control channel; green, valves; and red, sample-loading channel. Scale bar, 100 µm.**

## CENTRAL IDEAS UNDERLYING DEVICE OPERATION

The device designed by Kwanghun Chung was centered around the idea of using a single imaging channel and a simple scheme of “load, image, phenotype, and release/sort” (Figure 7a-d): first, a single animal is automatically loaded into the engineered microchip by a constant pressure-driven flow; then the animal is briefly (and reversibly) immobilized while multi-dimensional images of the animals are acquired on-chip; then phenotyping and sorting take place. The original device was designed to be either a two- or three-layer device.

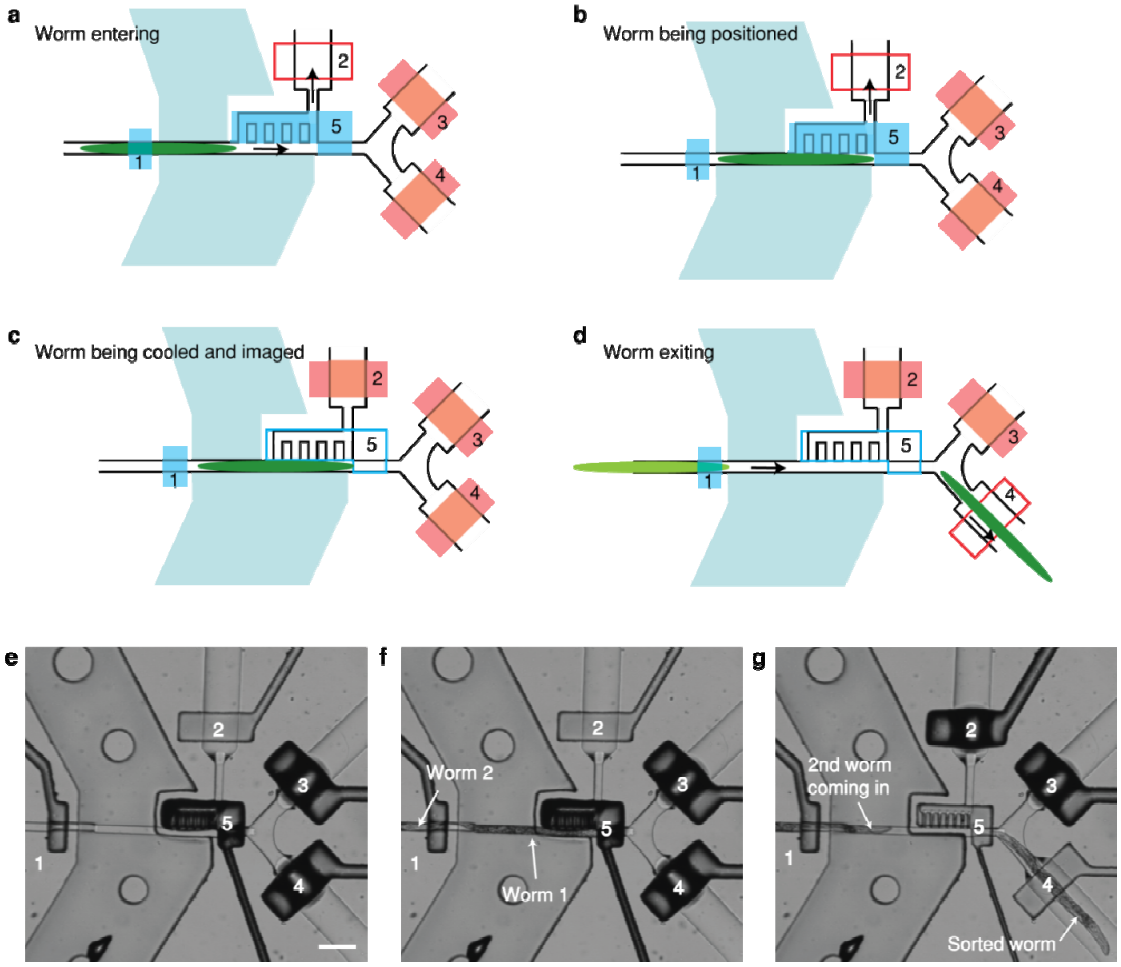



**Figure 7: Schematic diagrams summarizing the valve control sequence in the worm sorting process. Blue (dark): partially closable and tunable valves. Red (light): fully closable valves. Valve 1 is always partially closed to prevent multiple worms from entering. (a-b) Worm entering and being positioned: valve 2 on the positioning channel is opened to generate a pressure gradient to guide an animal into the observation chamber. (c) Worm being cooled and imaged: once the animal is positioned in the observation chamber all the valves for fluid to exit the observation chamber are closed to eliminate flow fluctuation. (d) Worm exiting: one of the exit valves (3 or 4) is opened to allow the imaged worm to leave. Once the worm leaves the observation chamber the valves return to the worm entering state. (e-g) Frames from videos showing positioning and routing of animals. Scale bar: 100 µm. (e) Waiting for an animal to enter. (f) A loaded animal preventing a second animal from entering. (g) The second animal is automatically moved into the detection zone after the previous animal exits the detection zone.**

The device performance relied on several engineered design features on the microchip. The chip has five salient features that ensure a consistent and reliable operation for an extended period of time. First, it automatically self-regulates the loading of nematodes by a simple passive loading-regulator design (Figure 8), as compared to using the complex multi-step loading mechanism that requires additional image acquisition, analysis, and valve actuation(59). Constant pressure drives the flow, so that no feedback or

intervention is necessary for the microchip to allow one and only one animal to occupy the imaging area at a time. Second, the setup automatically positions the nematodes in an identical position in the chip, so as to minimize the travel of the motorized stage and thereby reduce the processing time and increase the throughput. This is achieved naturally by the position of the animal relative to the positioning channels: when no animal is loaded, the positioning channels have a larger pressure drop and therefore larger flow, dragging the animal towards the far end; when an animal is loaded, the pressure force equalizes and the animal no longer experiences a significant force (Figure 8). By designing self-regulating loading and positioning mechanisms, we are able to reduce the complexity of software control and thereby significantly increase throughput and ensure that no time is spent moving the microscope stage to locate the animal.

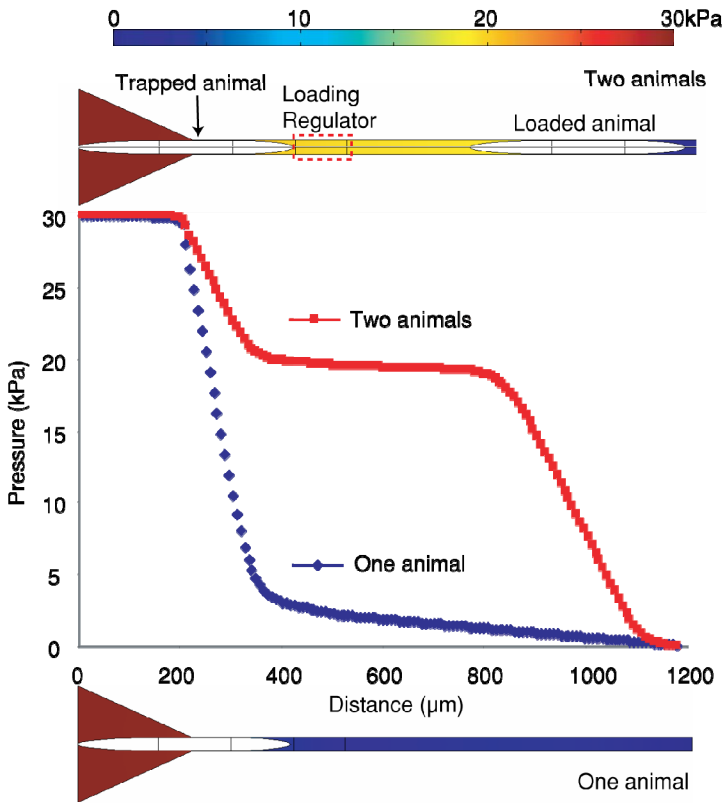



**Figure 8: Pressure profile in the imaging chamber determined by 3-D numerical simulation in COMSOL®. Top: top view of a two-animal model, showing pressure profile for a loaded animal in the imaging zone and a second animal waiting behind the loading-regulator.**

Third, the device has an integrated local temperature control system whereby animals are cooled to ~3-4 °C and completely immobilized briefly for imaging and manipulation without the use of anesthetic drugs. While mechanical immobilization of the animals merely by applying suction was found to reduce movement, contrary to what was reported earlier(59), it proved inadequate for high magnification imaging/sorting. The use of brief cooling, however, proved very effective at stopping the motion of animals and the GFP pattern of cooled animals compared to those immobilized with sodium azide showed no discernible differences. Thus cooling provides an alternative to anesthetics, a common practice in manual microscopy and laser ablation, potentially minimizing the adverse developmental effects. Fourth, the microchip has no permanent small features (<20 μm), and therefore is less easily clogged by debris and can operate relatively robustly. Lastly, losses through this system are minimal (~3%), and the viability of all the sorted animals is ~100%.

## PROBLEMS WITH THE ORIGINAL DEVICE

Although the device was used successfully to sort a few thousand animals, it had numerous weaknesses that led to a high failure rate during fabrication, inconsistent operation and rendered it incapable of performing continuous automated sorting.

The first problem was the difficulty of fabrication. The device created by Kwanghun Chung required a two-layer fabrication process that relied on plasma-bonding the two-layers of PDMS as opposed to thermal bonding. This process (shown in Figure 9), required cutting the thin film of PDMS that was spun coat onto around the perimeter of each device using a scalpel. This ~80 micron film was then flipped over a slide glass. The surface of the thick layer was then activated using the plasma, aligned and bonded. Because activation of the PDMS surface is a short term process, the alignment under a microscope had to be performed within a minute or the bonding might fail at the high pressures required to close the valves. Additionally, the bonding between layers was irreversible, so if a mistake in alignment was made, the

layers couldn't be separated to realign and bond. This, coupled with the short amount of time to perform alignment resulted in >40% of devices failing due to fabrication problems.

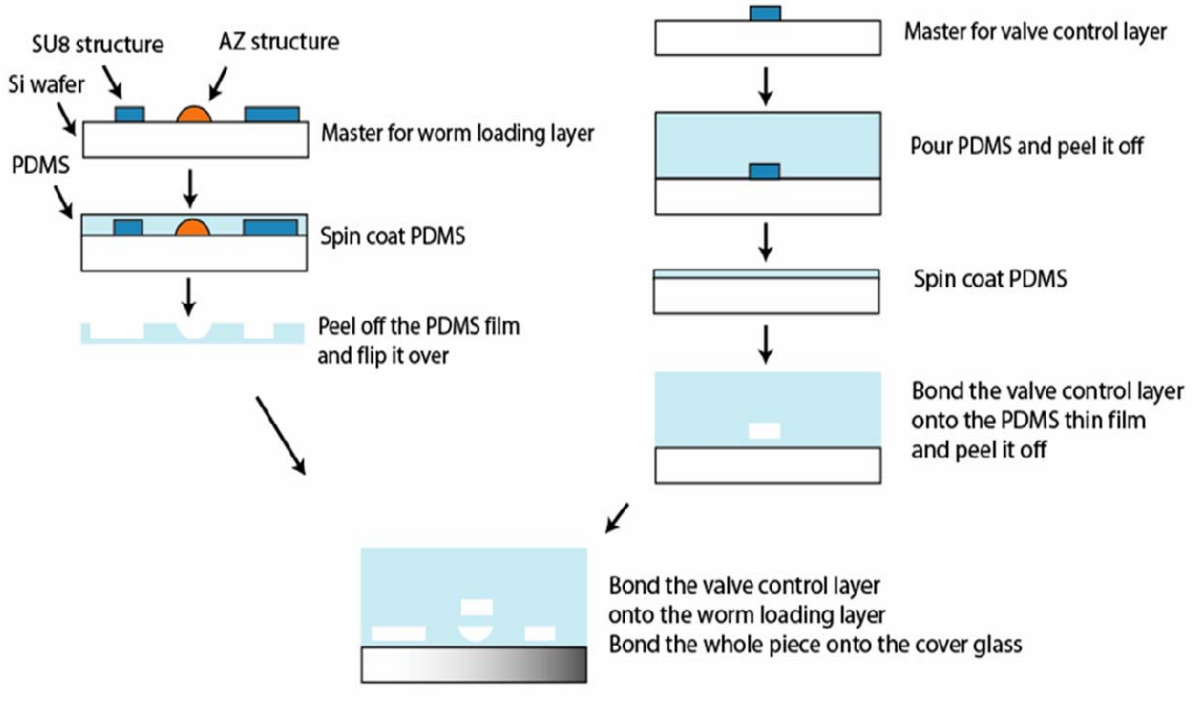



**Figure 9: The original fabrication method used in the lab. This process was extremely challenging and resulted in an unacceptably high failure rate.**

The second problem was the use of suction channels, and the problems they caused in design, fabrication and operation. The original idea underlying the suction channels was to guide the animals into the imaging region in a reliable and repeatable manner. Once the animal was positioned, the hydrodynamic resistance would increase, and flow would be reduced. The design problem with this was that it significantly restricted the dimensions of the coolant channel, because the channel had to wrap around the valves closing the suction channels. This caused uneven cooling across the animal that could be overcome, but created significant sensitivity to any fluctuations of the coolant fluid. This resulted in variations of up to 7°C across the length the animal(66). In order to immobilize the animal, with this device, it required that some of the animal be exposed to near or below freezing temperatures. This is

non-ideal, and could result in damage to the animal. Additionally, the placement of the cooling channel, the valves for the suction channels, and the actual imaging channel of the worm placed extremely tight tolerances on the alignment. Good device operation required alignment of +/- 10 microns, which is quite difficult to do by hand under a stereomicroscope. This further reduced the yield. Finally, as with any microfluidic device, there is the potential for debris and clogging. Although the suction channels were designed to be larger than those of competing devices(59, 61), they were still on the order or 20 microns in width, and more susceptible to clogging than other portions of the device.

The third problem was the inconsistent operation of the device depending on the density of animals in the tubing and device. Combining the fact that the worms are denser than water, and a Poisson distribution, there is a strong tendency for the density of animals in the device to increase or decrease during a sorting experiment that lasts several hours. When the concentration of animals in the main chamber, upstream from the imaging region, increases, the pressure drop across that portion of the device also increases. This in turn significantly lowers the flow rate through the device, and how the device behaves. Irregular flow rates cause significant problems because the field of view when using a high magnification objective is so small. When attempting to sort an animal to either as either mutant or wild-type, the device needs to perform consistently when sorting. The fluctuations in flow rate cause sorting to take as little as <1/2, or as long as 5 seconds. Furthermore, the only way to sort an animal is to allow unencumbered flow through the device, which can have the unintended consequence of allowing an animal behind the worm of interest to be sorted. Additionally, dealing with mutagenized populations means that there will be a large variation in animal size. These variations in device behavior made it difficult to create a system capable of operating for hours without human intervention or failure.

The fourth problem was the passive animal loading mechanism, and its inability to operate in conditions that were anything less than ideal. The idea of the passive loading was to use the deformable worm cuticle, and the hydrodynamic resistance change based on the presence of the worm to passively load

animals into the imaging region. If no worm was present, the pressure drop would drive a single worm into the imagine channel, and if a worm was present, the pressure drop would be too low to drive a second worm into the imaging chamber. This system, however, was susceptible to all the problems previously mentioned. If the density of the worms in the main chamber increased, the pressure drop across that chamber would increase, and worms would never be driven into the imaging chamber. Likewise, any variation in animal size would cause the pressure drop distribution to change and either prevent animals from entering the imaging channel, or cause them to double up. Even when using a non-mutagenized population of animals (thus the size distribution is tighter), the loading regulator would only work ~60% of the time.

## COMPUTER VISION

The fundamental idea of computer vision is to use statistical methods to learn specific characteristics about a specific image, or set of images, so that images with similar characteristics can be identified. Similar to how our brains extract specific shapes or use Gestalt perceptions, to extract information and understand what a specific set of colors means, computer vision aims to create a computational framework to allow similar high level understanding of images. The two central elements of computer vision are the extraction of specific quantitative features from the image, and the use of a specific statistical method to classify these features.

Formalization of the computer vision field began in the late 1970's, and continued to grow as computers became increasingly powerful. One of the earliest applications of computer vision was to the study of facial recognition(67-74). An early, and elegant, demonstration of facial recognition was performed by Turk and Pentland(68) in 1991 using the idea of eigenfaces. Using a series of face images, an eigenface space was created by determining the principal components of the sequence of images. To identify a new facial image, and determine whose face it most closely resembled, the new image was projected into the

eigenface space. The image in the database closest to the new face in the feature space was hypothesized to be the same person.

The development of computer vision methods has proliferated significantly since the eigenfaces example. Research has broadened from trying to classify individual faces, to identifying faces within whole images, gathering a semantic understanding of a scene, and even classifying medical images(73-84). Although the field has had significant growth, even to impacting commercially available systems such as those available from Google or Facebook, there are significant areas that could benefit from further work. One significant area is that of biological and biomedical research. Biological research generates massive amounts of data, and much of the annotation of the data (such as tracking cell movement) is still in most cases performed manually by highly trained experts. Automation of these tasks would not only free up resources, but would allow large-scale systems biology experiments to be performed that are currently not feasible solely because the time required to analyze the data would be prohibitive(79, 80, 85, 86).

## MACHINE LEARNING METHODS

Currently there are a large number of different statistical methods that can be used in computer vision. Developing novel methods or tweaking old methods for slightly faster of more accurate performance, is a significant research field in its own regard. The following list is merely meant to act a simple primer in some of the more popular current methods, but will likely soon become outdated as newer methods become in vogue. Although each method has specific advantages and disadvantages, as long as good procedures (such as normalization and cross-validation) are followed, any of the following methods will perform adequately for most computer vision needs.

### SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) are classification method based around the idea of identifying the hyperplane that best separates two sets of points. The underlying theory was developed in the 1960's, but it was only in the late 90's that they become popular in a machine learning context. The simplest way of understanding SVMs is that they attempt to find a line that best separates to sets of points, and maximizes the margin separating the classes on each side of the line (hyperplane)(87). The name is derived from the points in each class that are located along the margin and called support vectors.

The general form of SVM is(87):

$$g(x) = \beta_0 + \sum_{i=1}^{N} \alpha_i y_i \langle x, x_i \rangle$$

Where the classification rule is $sgn(g(x))$ and SVM optimization requires the minimization of

$$\sum_{i=1}^{N} \left(1 - y_i g(x_i)\right) + \lambda \|\beta\|^2$$

This SVM formulation is a linear classifier, but can be easily transformed into a non-linear classifier using the kernel trick(88). This allows the data to be mapped into a higher order space, where separation is often performed better, and thus allows significantly increased classification performance. Because the formulation uses a dot product, it can be replaced by any kernel that is positive and semi-definite(89). Several common kernels are frequently used, and in this work we used the Gaussian radial basis function (RBF) kernel when not using a linear classifier. This is formalized as:

$$k(x, x_i) = e^{(-\gamma \|x - x_i\|^2)}$$

$$g(x) = \beta_0 + \sum_{i=1}^{N} \alpha_i y_i k(x, x_i)$$

One of the advantages of using support vector machines for pattern recognition applications is the large number of curated libraries. Rather than requiring users to reinvent the wheel, these libraries are cutting edge implementations that are extremely fast. Among the different libraries for SVM use are: SVM-light, libsvm, liblinear, MLPack, flssvm, Shark, dlib, and the Matlab default. Many of these libraries were written for specific languages, but have been ported so that they can be used with most programming languages. For this work we used both the libsvm(90) and liblinear(91) libraries, compiled and then accessed from Matlab. They were selected based on speed and ease of use.

## BOOSTING/BAGGING

Boosting is fundamentally the idea of combining a number of weak learners into a single strong classifier, and thus dramatically increases the classification strength, even when using classifiers that individually would perform extremely poor(78, 80, 92-95). For training purposes each weak learner is trained, and then the minimum classification error is computed to determine the classifier with the lowest error. This classifier is based on the performance (classification error), and then the process is performed again to determine which classifier helps to reduce the error the most given the classifiers that have already been weighted and trained. During the weighting and retraining process, examples that have been misclassified are weighted higher, whereas correctly classified examples lose weight(92). Repeating this procedure allows the strong classifier to be gradually trained and approach the best combination of weak learners(96). In some cases boosting has been found to be less sensitive to overfitting relative to other machine learning methods. Because the primary point is to combine weak learners, there are a large number of different boosting methods that use different error functions and weighting methodology. Several of the most popular boosting methods are AdaBoost, GentleBoost or LPBoost(97-100).

The general function for boosting classifiers follows the form(92, 96):

$$f(x) = \sum_{j=1}^{J} \alpha_j h_j(x)$$

Where $\alpha_j$are weightings each of the weak classifiers, and $h_j(x)$ is the output of the weak classifier $[-1, 1]$. The greater the$|f(x)|$, the higher the confidence that the instance x is part of the particular class. Although the Gentleboost algorithm was originally implemented and applied to binary classification, it can also be applied to multiclass situations.

## MULTICLASS TRAINING METHODS

The default design for SVM and boosting is two optimize binary classification schemes (something is either A or B). In many situations, however, it is necessary to classify and object or image as being one of potentially many classes. This is commonly resolved by converting the classification into a sequence of binary classifications. The method used in this work was the one-vs-all approach. In this, each classifier is trained against all of the other classes. When a test point is introduced, all of these binary classifiers are applied to the new point, and the class that gave the largest result is selected.

## FEATURE EXTRACTION METHODS

Because of the large amount of interest both academically and commercially, for computer vision tools that are capable of high levels of performance, there a number of different feature descriptors. These descriptors extract specific features from an image, or sub-window within the image, to create a quantitative framework that can be used for computer vision. Once extracted, these quantitative features are used in conjunction with the machine learning algorithms to detect specific patterns in images. Most descriptors were developed for use in everyday situations to detect people, cars, or other conventional objects. As discussed in later chapters, certain features were adapted and modified to perform well with fluorescent images in a biological context.

## HISTOGRAM OF GRADIENTS

The histogram of oriented gradients (HOG) is one of the commonly used feature descriptors for computer vision and object detection. It was originally developed by Dalal and Triggs(101) and is similar in some regards to the SIFT features described next. The fundamental idea behind the HOG descriptors is that within a window, both the magnitude of the gradient and the orientation provides a large amount of information about what is contained within the window. The standard HOG implementation separates the image into smaller windows or cells in which the descriptors are calculated. Within each of these windows, the gradient is computed – typically with a [-1 0 1] and $[-1\ 0\ 1]^T$ matrices. The gradient is then binned based on the orientation of the gradient. Within this histogram, the magnitude of the gradient in each direction is used. The histogram is either "unsigned" and goes from 0 to $\pi$, or is "signed" and goes from $-\pi$ to $\pi$. These cells can be relatively small sub-windows of from the larger image, and are typically combined into larger blocks to allow for semantic understanding of each cell in relationship to each other and the image as a whole. Combining these smaller cells into blocks is typically done by either taking rectangular numbers of blocks (R-HOG) or a circular number around the center block (C-HOG)(101).

## SCALE INVARIANT FEATURE TRANSFORM

The Scale Invariant Feature Transform (SIFT) features are one of the more popular feature descriptors used in analyzing images and computer vision. The SIFT features were created by Lowe in 1999(102), and have been applied extensively to computer vision problems. The SIFT features were created out of difficulties of obtaining good quantitative descriptions of objections when the object was distorted, for example if the object was scaled because it was further from or closer to the camera. The feature descriptors are designed to be invariant to scaling, rotation and image translation.

Rather than performing a dense feature extraction on an image like HOG, SIFT identifies specific keypoints that likely contain more information about the image than surrounding regions. This is done in a four step process(103)

1) Scale-Space extrema detection. The original image is convolved with Gaussian filters at different scales and the difference between the filtered outputs is computed. Maxima/minima of these Difference of Gaussian (DoG) images are identified as potential keypoints.
2) Keypoint Localization. Using the previously identified points, they points that appear to be less stable of different scales are removed.
3) Orientation assignment. Keypoints are assigned an orientation assignment based on the local gradient. Future operations compare objects transformed relative to these gradients to minimize variance with regard to orientation.
4) Keypoint descriptor. The local gradients around the keypoints are used to create the feature set that is then used for comparison. Because the features are transformed, they allow for significant distortion in local shape and illumination change.

## BAG OF WORDS

The idea underlying the bag of words (BoW) method is that invidual descriptors mean more when put in context to the descriptors around it(104-108). This is analogous to the idea that by analyzing the frequency of specific words in a document, you could predict with some degree of accuracy what was being conveyed by the document. The BoW idea is often used for object or image categorization, such as identifying whether a particular image is of a city, mountain, ocean, etc. Rather than relying on individual descriptors, it was found that in some cases, using the descriptors to predict what each sub-window an image described (word), and then using the frequency of the words to describe the image as a whole could increase the classification accuracy. Rather than developing a new feature set, the BoW uses

quantitative feature descriptors (such as HOG or SIFT) to extract quantitative information and to predict what "word" is displayed in the sub-window. Analyzing the frequency of words within an image then allows classification of the image into the specific category.

The generation of the "words" used in representing images can be generated either by expert labeling of images to group similar objects or by clustering of all the regions to identify objects that are more likely to be grouped together. Clustering by k-means, or similar method, is the most common way to generate the words. Undirected clustering allows the creation of a large dictionary of "words" with minimal labor and avoiding potential human bias. In generating this dictionary, however, it is important to have a large database of images for training purposes so that small changes in scale, lighting, or other situational differences don't result in poor clustering. Although BoW has been demonstrated to work extremely well in many situations, one of the notable drawbacks to this method is that it merely uses the frequency of the individual words for prediction and not the relative positions or contexts.

## MACHINE LEARNING AND VISION APPLIED TO *C. ELEGANS*

Although there have been numerous high-profile applications of machine learning methods applied to high-content imaging and analysis of biological experiments, it has been primarily limited to cell or tissue cultures(27-29, 32, 79, 80, 85, 109). In many of these examples vision methods were used to identify specific features in the cells, and then used to extract novel information based on the high-content information extracted. Information extracted could range from cell morphology(27) to colocalization of fluorescent markers(110). Applications to multicellular model organisms, however, have largely been limited to identifying movements or the analysis of complex behaviors(83, 84). Within the C. elegans community, statistical methods have primarily been applied to reduce the phenotypical space to cluster animals based on specific extracted features(111-113). Some work has been performed on lineage tracking during early embroygenesis that used machine learning and SVMs to increase the classification

and tracking accuracy(81, 114-117). All of these examples, however, are significantly different from this work as information was gathered and then machine learning tools were used to analyze the experimental results.

## EXTERNAL CONTROL COMPONENTS FOR MICROFLUIDIC SYSTEMS AND LACK THEREOF

Systemic, or off-chip, components are central to the robust operations of microfluidic chips, although this is often less emphasized in microfluidic literature. Well engineered off-chip components are critical to creating an ease of operation sufficient to allow microfluidic devices to tackle biological problems of interest. The following are important areas to consider when designing a microfluidic system: image analysis and signal processing, automation (through error handling and valve control), incorporation of appropriate microsurgery laser tools, and lab-to-chip interface methods. There are currently an abundance of biological problems that require significant advances in modern technology before being solved, but without the creation of a comprehensive system solution, microfluidic devices would fail to have the significant impact on the biological community that could otherwise be achieved. Furthermore, because the microfluidic field is a recent development, primarily of academic interest, commercial support for systemic components is extremely limited.

Autonomous, or semi-autonomous operation requires devices to be integrated with computer controls(1, 58). Typically this entails off-chip solenoid valves actuated by a digital control board, electronically controlled pressure regulators, a computer controlled image acquisition system, and potentially an x-y-z stage. Automation, however, requires significantly more than computer control. The size of a multicellular organism can vary dramatically, and that coupled with motility and the ever present problem of clogging by debris necessitates extensive error-handling. Creating a fully automated system requires a collaborative effort with computer science, electronics, and robotics. Although the

requirements for automated microfluidic systems used with multicellular organisms differs little from those used with particles or single-cells, commercially available systems that integrate with microfluidics are not available at present, and custom-made systems are typically used.

Microfluidic systems that meet some of the highest impact biological needs, such as laser microsurgery or microinjection, also require additional specialized hardware. Laser microsurgery on model organisms has been extremely useful since it was first popularized for killing specific cells(118) and studying development. Using a low-powered, nano-second laser allows researchers to selectively ablate cells and study the behavioral response when a lesion is created in the network. Taking advantage of the precise control and higher throughput, this has recently been demonstrated using an automated microfluidic system(57). Additionally, the utility of femtosecond lasers (although considerably more expensive) has allowed researchers to dramatically limit the amount of energy absorbed by surrounding tissues and to ablate axons or dendrites to study axonal regeneration. Two groups have demonstrated the ability to immobilize *C. elegans* in a microfluidic device and cut axons (60, 61, 119). One group reported that the use of microfluidic immobilization resulted in significantly more rapid axonal regeneration when compared to conventional methods using anesthetics(60). Another area of interest is the injection of *Drosophila* and zebrafish embryos with RNAi for large-scale forward genetic screens(120-124). Using microneedles developed by conventional silicon MEMS fabrication coupled with a microfluidic device, the ability to inject large numbers of embryos with RNAi has been demonstrated.

Most microfluidic devices are currently created and operated in engineering labs where the difficulty of setting up and replacing devices is secondary to creating novel designs. As such, most microfluidic devices are painstakingly set up for each experiment by connecting small pieces of tubing to holes cored into the PDMS device. Not only is this a time consuming bottleneck, but it creates a failure point and limits the ability to use standard lab equipment such as micro-pipettes or liquid-handlers in conjunction with the microfluidic systems created. In the near future, one could expect that interfaces capable of meeting the

needs of many labs and applications can be developed and commercially marketed. This would not only speed up the development time for new microfluidic systems, but reduce the early adopter cost to biology labs interested in utilizing systems already published.

The original external control systems present in the Lu lab were minimal, and none of them allowed for computer control. Controlling the on-chip micro-valves required manually turning a switch on or off to control the pressure lines powering each individual valve. This method would be sufficient for device prototyping, but not for automated imaging and screening. Furthermore, image acquisition and analysis was performed using proprietary software. Creation of an integrated closed loop system required the development of specific external components that could use the acquired images to determine the appropriate valve sequence.

## THESIS OBJECTIVE AND CONTRIBUTIONS

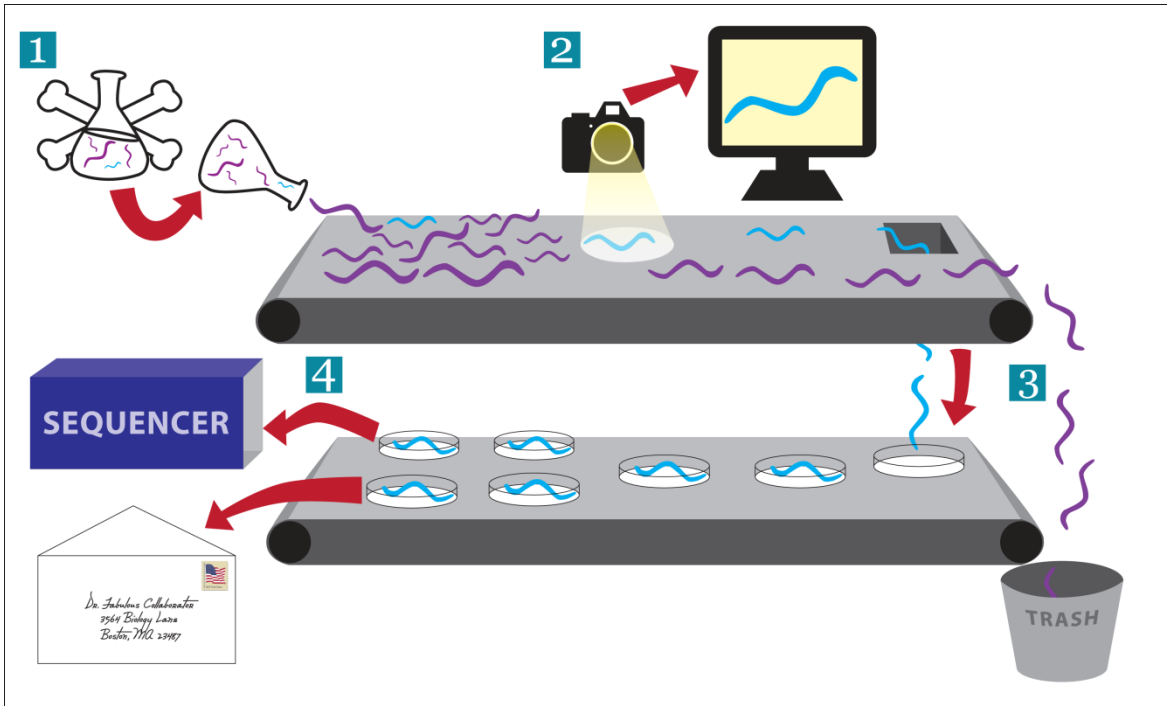



**Figure 10: System vision: an automated screening system that would be capable of discovering novel phenotypes and sorting thousands of animals without human intervention. A mutagenized population of animals (1) is introduced to a system that automatically images and phenotypes (2) the animals to identify specific features of interest that allow sorting (3); the sorted animals are then sequenced and sent to collaborators (4).**

This thesis makes contributions to the study of neurobiology through the creation of a system for phenotyping and screening the model organism *C. elegans*. The major intellectual contributions are:

1) The creation of an optimized microfluidic device that is capable of robust handling of *C. elegans* to enable imaging and sorting at high-magnification. In contrast to previous systems, this device has been designed so that it performs consistently and is capable to operating continuously.
2) Creation of external systems to allow computerized control and error handling of the microfluidic device. This enabled continuous operation without human intervention.

The development of a computer vision framework to identify subtle fluorescently labeled features (synapses, cell-bodies) and to distinguish them from surrounding tissues and autofluorescence. This

computer vision framework uses state-of-the-art algorithms to identify the objects of interest in an extremely poor signal-to-noise environment.

3) The creation of quantitative features that allow for the creation of a quantitative phenospace significantly more advanced than alternative phenotyping methodologies in developmental neurobiology.
4) Demonstrating the utility and power of the combined system by performing large-scale forward genetic screens to identify novel genes involved in synapse formation and guidance. Completely autonomous screens have identified several classes of animals previously unknown. These were the first genetic screens performed on a multicellular model organism using a microfluidic system.

# CHAPTER 2: MICROFLUIDIC DEVICE DEVELOPMENT FOR HIGH-THROUGHPUT SCREENING

The original generation of microfluidic devices contained numerous limitations that limited the effectiveness in performing large-scale high-throughput screens. To overcome the design flaws, discussed at length in Chapter 1, a microfluidic system was created that would be capable of extended operation and could be used for high-throughput, automated screening. This was done be redesigning the first generation device to reduce failure modes and to reduce fabrication complexity.

## REDUCING FAILURE RATE BY PARING FEATURES AND MICROFLUIDIC REDESIGN

The original system designed by Kwanghun Chung had a significant number of features that, while elegant in design, proved incapable of performing robustly for extended operation. Additionally, the design of the device and reliance on fully closable valves made the fabrication complicated and resulted in a high failure rate. This made performing actual experiments with it extremely challenging, and made transferring it to biology labs extremely impractical. The problems encountered while using the device designed by Kwanghun Chung are summarized below in Table 1.

**Table 1: Summary of the problems encountered using the original device design.**

| Problem | Details |
|---|---|
| Difficulty of fabrication (low fabrication yield) | Suction channels require +/- 10 micron alignment<br>Aligning two-PDMS layers in <1 minute because of plasma bonding<br>Peeling off thin-film and inverting it before alignment |
| Suction channels design | Small channel dimensions result increased clogging relative to rest of device<br>Uneven cooling of the worm in most devices due to small mis-alignment of cooling channel |
| Inconsistent device operation with varied worm density | Device has only a single pressure and fluid source. Any change in resistance in the device changes the system operation dramatically. |
| Loading regulator works only in idealized situations | Passive loading regulator relies on identically sized worms, and consistent device pressure drop. Any change in either, and it fails to work |
| Temperature gradient across animal | The design of the cooling channel creates a large temperature gradient across the animal. |

In order to address the various issues presented above, and in greater detail in Chapter 1, the device components were pared back to a much simpler version. Rather than attempting to create a device that could be used to perform automated screening in a single attempt, the objective was to optimize all of these components to create a device that could be used for rapid visual screening. The following design objectives were chosen of the device:

1) Ease of fabrication. The devices should be designed in such a manner that a trained undergraduate student would be able to have a device yield of >90%. This would make the device transferrable to collaborators and significantly reduce the labor requirement of doing a large scale screen.
2) Resistance to clogging. The device needs to have a minimal number of small features that could be clogged so that it could be used for an extended period of time before failing.
3) Ease of setup. Each additional on-chip valve not only increases the failure rate of during the fabrication, but the setup time and chance of failure during setup. The device should require the absolute minimum number of valves and fluid entrances/exits.

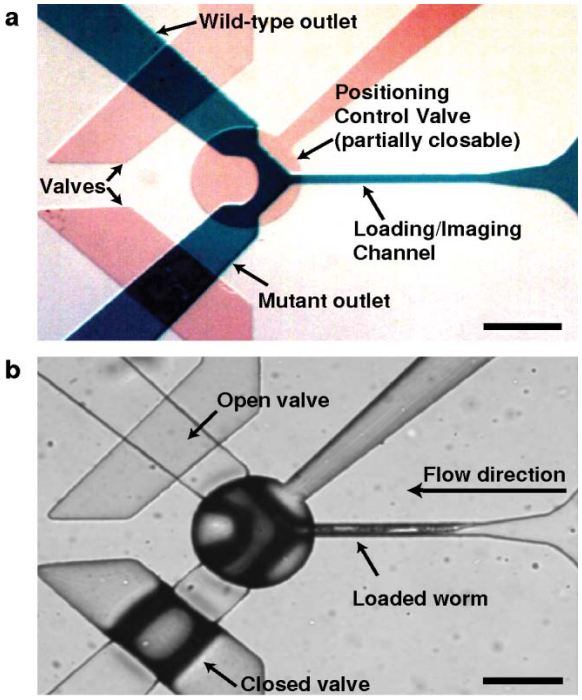



**Figure 11: Microfluidic device for rapid screening (a) Optical micrograph of the device active region. Channels are filled with dyes to reveal key features: blue, sample flow layer; red, valves. (b) A frame from a video showing a worm being loaded into the field of view.**

The completed device is shown in Figure 11. When the circular valve is pressurized, it only closes partially allowing fluid flow to continue but creating a cross-section too small for the animals to pass. At the same time, the valve controlling the wild-type outlet is left open most of the time to allow fluid to exit in order to load animals for rapid processing. This mechanism reduces the number of valve operations since mutants are rare events. To further increase throughput and simplify the sample-handling procedure, animals were not deliberately immobilized by cooling or pressure, as screening with a 20x objective with a relatively bright marker does not require complete immobilization. To reduce the likelihood of another worm entering the field of view when one is already loaded, we designed the device geometry such that the presence of a worm in the imaging channel increases the resistance of flow significantly and thus reducing the flow by an order of magnitude.

## TURNING DRAWBACKS INTO ADVANTAGES: USING PARTIAL CLOSURE VALVES

One of the central elements of the original design was its use of on-chip valves that could be completely closed. These valves close completely when pressurized because they have a rounded cross-section, and it was the need for fully closed valves that significantly complicated the fabrication process. Eliminating fully closable valves was a requirement for simplifying the fabrication procedure and increasing the yield rate. Instead of the complicated fabrication procedure shown in Figure 9 that Kwanghun Chung devised, this allowed the device to be fabricated using conventional thermal bonding between layers (as seen in Figure 62).

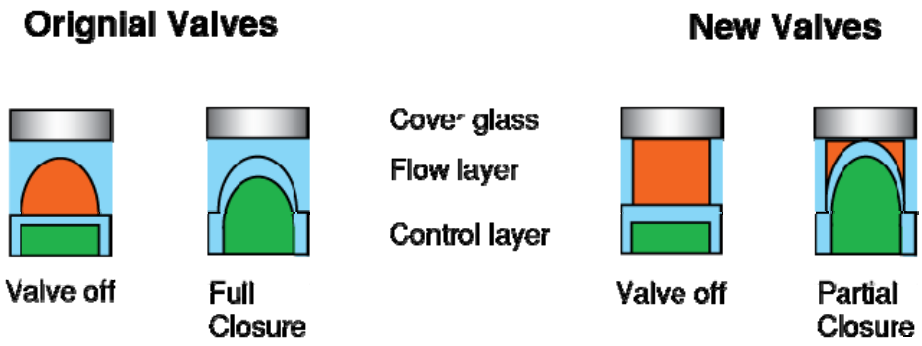



**Figure 12: Schematic showing the original full closure valves and the new, partial closure valves. Using partial closure valves significantly simplified the fabrication procedure and reduced the failure rate.**

Although valves incapable of fully closing are typically considered drawbacks in microfluidic devices, in this case they could be used advantageously to simplify the design considerably. Because the valves always allow a small amount of flow, the new device was designed using partial closure valves to position animals in the imaging channels, instead of the suction channels. The new valves and the comparison to the original design are shown in Figure 12.

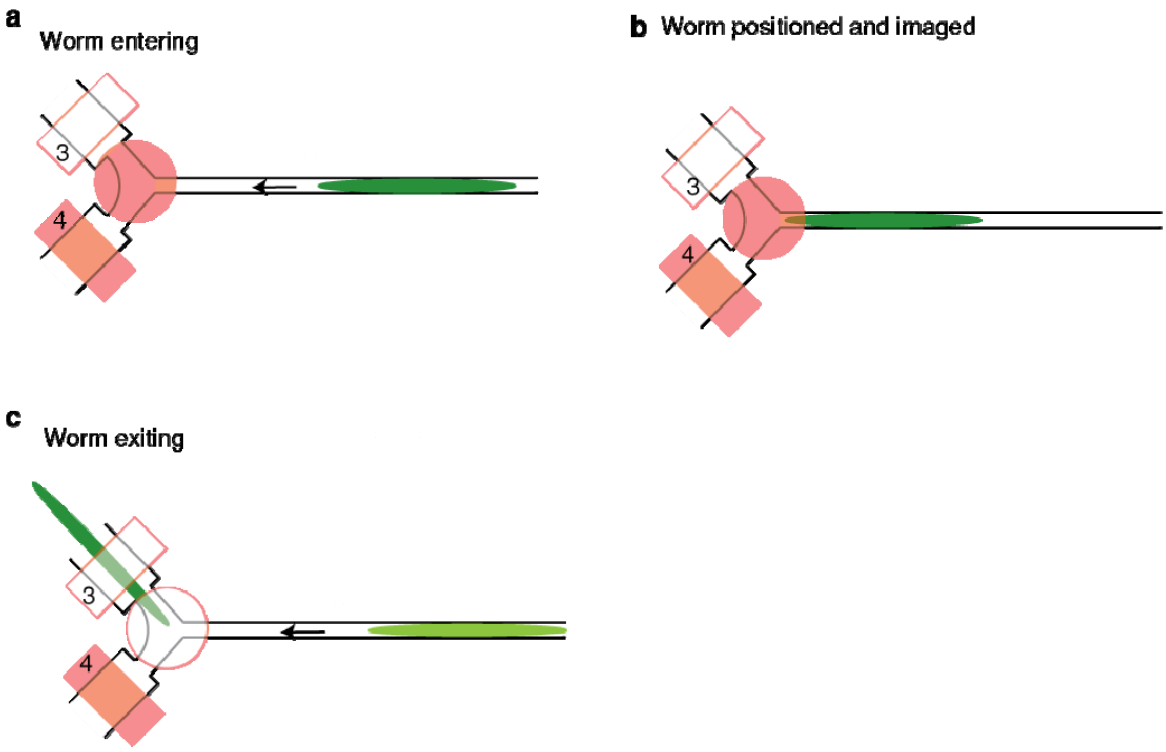



Figure 13: Schematic showing the sequence of steps for device operation.

During the loading process (seen in Figure 13), the openings near the wall underneath positioning control valve allow moderate fluid flow through the device. This serves to push the worm through the device and into the imaging channel. Once a worm is pulled into the imaging channel, the flow pushes it up against the position control valve, which stops it. Because the gaps under the valve are substantially smaller than the worm, it is unable to move further forward, and is stopped. At this point the presence of the worm within the imaging channel causes a significant change in the fluidic resistance, and creates a corresponding decrease in the flow rate. A COMSOL fluid dynamics model was generated to show the effect of a worm on the flow and pressure profiles within the device. As can be seen from in Figure 14, the flow decreases dramatically in the presence of an animal.

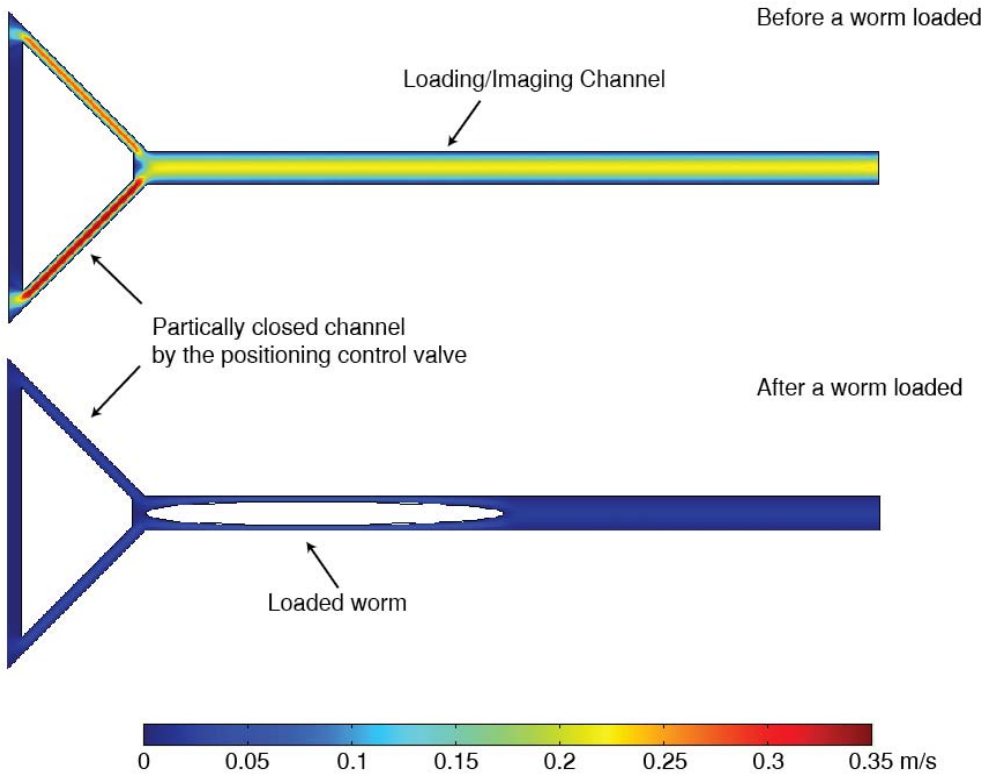



**Figure 14: Numerical models showing flow rate distribution along the worm loading channel before and after a worm is loaded. To simplify the numerical simulations, partially closed channel by the positioning control valve is assumed as channels with a small width (10 ~ 12 μm) along the edge of the original channel geometry. The presence of a worm causes at least 10-fold decrease in average flow rate: 0.153 m/s before a worm loaded, 0.0137 m/s after a worm loaded(1).**

## DECREASED FAILURE MODES: FABRICATION AND CLOGGING

This redesign to use continuous flow allowed me to design around several problems that plagued the original, especially in regards to failure modes during both fabrication and operation. The original fabrication was extremely complex, resulted in a low yield, and as previously discussed resulted in a fabrication failure rate of nearly 50%. Modifying the design by using partial closure valves allowed for the simpler fabrication process with thermal bonding between layers, and reduced the fabrication failure by half.

Making a device that would be transferable to other labs, however, would require dramatic increases in fabrication yield to a point where large numbers of devices could be easily made by students relatively inexperienced in microfluidcs. To meet this objective and further enhance the fabrication yield, the device was designed so that it has a larger tolerance to minimize the consequences of poor feature registration (either rotational or translational). As can be seen in the images of the device, the valves appear oversized relative to the device. Whereas the first generation of devices required alignment tolerances of +/- 10 microns and a rotational alignment of a few degrees, this device can perform well even with significant misalignment. Even when fabricated by undergraduate students spending a summer in the lab, device yields were >90% after a few weeks of training. This design principle has allowed the device to be easily duplicated by users unfamiliar with microfluidics and increased the impact of this research.

This design allows efficient and well-controlled loading of animals, even with size variations resulted from the mutagenesis. By eliminating the suction channels, the device minimum feature size of the device was increased to 30 μm. Because the smallest feature in a device is that which is most likely to cause a significant problem and result in clogging, this allowed extended device operation.

## COMPUTERIZED DEVICE CONTROL

To take advantage of the human flexibility and the computer quantitation, I developed a computer-assisted methodology using Matlab® to allow an expert to determine in real-time, whether animals being screened are of interest. Using this software control interface (Figure 15), preconfigured image processing modules can be selected if needed to help clarify and accentuate phenotypical characteristics. While animals clearly exhibiting no interesting phenotypes can be dismissed quickly, potential mutants can be examined in greater detail using the image processing modules on the same user-interface. When a worm is in the field of view, one of over forty combinations of image processing options can be selected and subtle phenotypes emphasized.

For markers that are out of focus, one option is to acquire a small z-stack of images at different focal planes (with user-determined step size and number), and either autofocus or flatten the z-stack before further processing the images. This significantly reduces the time relative to manual focusing of the microscope and searching for the reporter, and potentially avoiding photobleaching of the markers. There are eleven image-filtering options to accentuate features of interest, which tend to be dim or low in contrast to human eyes, but the phenotypes become more obvious with image enhancement.

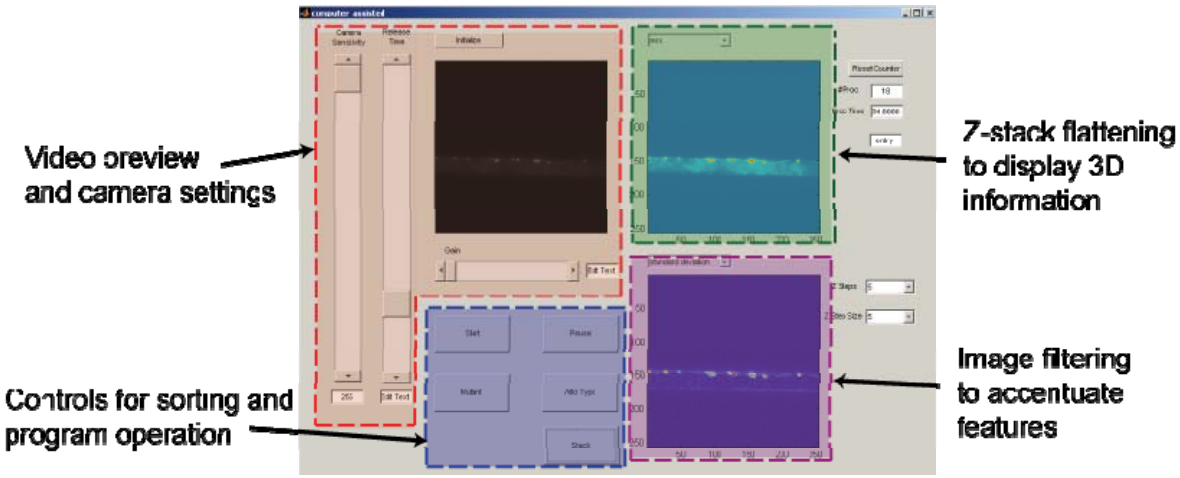



**Figure 15: The computer control interface. The video feed is shown in the top left box, and image processing steps can be selected, applied and displayed in the boxes on the right. Animals are sorted as either wild-type or mutant by selecting the appropriate button. If an image is unclear, pictures can be acquired at multiple focal planes and processed using selected image processing modules.**

The software interface written in Matlab® allows users to control various camera settings such as EM sensitivity and gain, and to control the exit time of animals. By selecting the appropriate buttons, an animal is sorted as either mutant or wild-type. If the image in the streaming video window is unclear to the user, selecting “stack” will acquire images at multiple focal planes (number of and spacing of images as specified by the user). Images can be flattened and processed according to the user selection. The following toolboxes are required for use: data acquisition, image acquisition and image processing.

The various options for flattening the z-stack are:

- **Summation**: this flattens the stack by making each x-y point equal to the summation of the values at that point over the z-direction
- **Maximum**: this flattens the stack by making each x-y point equal to the maximum of the values at that point over the z-direction.
- **Standard Deviation**: this flattens the stack by making each x-y point equal to the standard deviation of the values at that point over the z-direction.
- **In-focus**: this assumes that the slice with the highest standard deviation is the most in-focus and uses it for the subsequent image processing steps(125, 126).

The various options for image processing the flattened image are:

- **Gaussian**: applies a rotationally symmetric Gaussian low-pass filter to the flattened image.
- **Laplacian**: applies a filter approximating the Laplacian operator to the flattened image.
- **Laplacian of Gaussian**: applies a rotationally symmetric Laplacian of Gaussian filter to the flattened image.
- **Prewitt H**: applies the prewitt filter for emphasizing horizontal edges to the flattened image.
- **Prewitt V**: applies the prewitt filter for emphasizing vertical edges to the flattened image.
- **Sobel H**: applies the sobel filter for emphasizing horizontal edges to the flattened image.
- **Sobel V**: applies the sobel filter for emphasizing vertical edges to the flattened image.
- **Unsharp**: applies an unsharp filter for contrast enhancement created by the negative of a Laplacian filter and applies it the flattened image.
- **Range**: filters the image provided by the flattening step using the local range of the image.
- **Entropy**: filters the image provided by the flattening step using the local entropy of the image.
- **Standard deviation**: filters the image provided by the flattening step using the local standard deviation

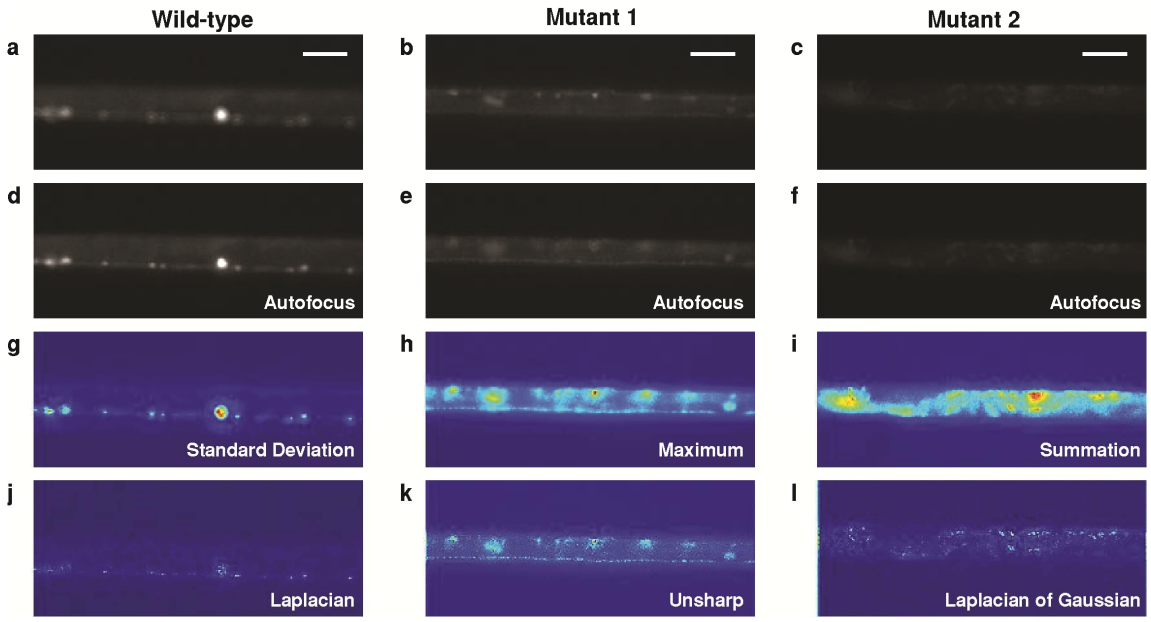



**Figure 16: Computer-assisted phenotyping to identify mutants of interest. (a,d,g,j) Wild type. (b,e,h,k) Apparent synaptic mutant showing altered reporter expression along the nerve cord and puncta structures. (c,f,i,l) Mutant showing reduced YFP expression. (a-c) Images of animals that entered, not necessarily in focus and potentially rotated, resulting in an unclear image of the region of interest. (d-f) Images determined to be in-focus by computer after a series of images at different focal planes was acquired. (g-i) Selected alternative methods of viewing z-stack by flattening the matrix of images. (g) Flattening by taking the standard deviation of the z-stack at each x-y location. (h) Flattening using the maximum value at each x-y location. (i) Flattening by taking the summation in the z-direction at each x-y location. (j-l) Applying a few of the image processing features to the flattened image to accentuate different features. (j) Laplacian filter. (k) Unsharp filter. (l) Laplacian of Gaussian filter. Scale bars 30 µm.**

## THE FIRST GENETIC SCREEN OF A MULTICELLULAR MODEL ORGANISM IN A MICROFLUIDIC DEVICE

Prior to this work, research using multicellular model organisms in a microfluidic device had been limited to proof of principle experiments that demonstrated the potential utility of a rapid imaging(59) or to engineer controlled microenvironments for more precise experimentation(47, 48). These experiments, while useful, had yet to demonstrate that microfluidic systems could be used to perform highthroughput screening experiments that would be both biologically interesting, and provide concrete results. In order to demonstrate the ability of microfluidics to meet this need, a forward genetic screen looking for differences in of localization of a fluorescent reporter was performed. The purpose of the screen was to identify novel genes involved in synapse vesicle trafficking and synapse formation. This requires

identifying novel mutations based on changes in fluorescent reporters that are localized at synapses, which is a challenging task as it requires a high NA objective. Screening throughput when done by hand is typically limited to only a hundred animals an hour. Screens for genes affecting these biological processes are thus far from saturation due to the degree of difficulty and labor required.

## SCREENING DETAILS

A forward genetic screen was performed in a successful screen of an ethyl-methanesulfonate (EMS)-mutagenized *C. elegans* population carrying a synaptic reporter punc-25-YFP::RAB-5(127). These reagents were provided by Dr. Yishi Jin from UCSD. The protein *rab-5* is typically located at the presynaptic terminals, and the purpose of the screen was to identify genes that when removed would lead to a change in the amount of the fluorescent reporter localized at the pre-synapse in GABAergic motor neurons. Because of this, it required a high magnification objective to see the changes in synapse expression. Synaptic vesicles are required for release of neurotransmitters and subsequent communication between neurons. The synaptic vesicle membrane proteins that facilitate fusion with the plasma membrane are transported from the cell body in synaptic vesicle precursors. These precursors must undergo a maturation process before they become functional synaptic vesicles, and this process likely involves fusion with endosome-like compartments within the synaptic terminal. The Rab-5 GTPase is associated with canonical early endosomes and is known to be present in presynaptic terminals. Previous research has shown that biasing the cycling state of Rab-5 towards a primarily GTP-bound form appears to lead to increased association of Rab-5 with endocytic membrane within the synaptic terminal, resulting in a concomitant decrease in the numbers of synaptic vesicles(128). For more details on procedure, see the Appendices.

Most animals can be processed in <2 seconds, and sorting was achieved at rates up to 2500 animals per hour and at a sustained rate of at least 1500 per hour.  From a screen of ~15,000 mutagenized worms, a

number of novel mutants were identified that appeared different from wild-type (Figure 16). The identification of these mutants was greatly facilitated by applying the auto-z-stack and autofocusing (Figure 16 d-f) and the many filters (Figure 16g-l). One class of mutants involved changes in the localization of the reporter yellow fluorescent protein (YFP) localization along the nerve cord and significant numbers of punctated structures (Figure 16b,e,h,k), and another class showed a dramatically reduced expression of YFP (Figure 16c,f,i,l). Some of the mutants would have been difficult to identify using conventional methods.

## EVALUATING DEVICE PERFORMANCE DURING THE LARGE-SCALE SCREENING

The primary objective of this screen was to demonstrate the ability to perform large scale screens using computer software to control the device, but utilizing human decision making in classifying animals as either mutant or wild-type. This proved successful as the high-throughput and the rapid sorting allowed the analysis of thousands of animals within a few hours. The secondary objective of this, however, was to fine-tune the device and assess the strengths and weaknesses in order to design a third-generation system that would be capable of performing fully automated screens.

### LOADING REGULATION

The loading regulation proved sufficient for a rapid visual screen. Insuring that only a single animal was present was primarily done by having a Poisson distribution of arrival events, and the decreased flow rate that resulted from the presence of a single worm. During preparation for the screening, the density of animals injected into the device was carefully controlled, and kept below ~2,000 animals per ml. Above this concentration of animals, the presence of two worms within the imaging channel increased significantly. The reduced flow rate (by an order of magnitude) when a single worm is present, shifts the arrival distribution to the significantly longer times, typically ~4-5 seconds. If more was required for

decision making, the input pressure was turned off, and this provided substantially more time before an additional worm appears. Depending on the skill of the user performing the analysis, this is sufficient to allow decision making and processing of the animals. In Figure 18, below, you can see the distribution of the processing times during one of the actual screens.

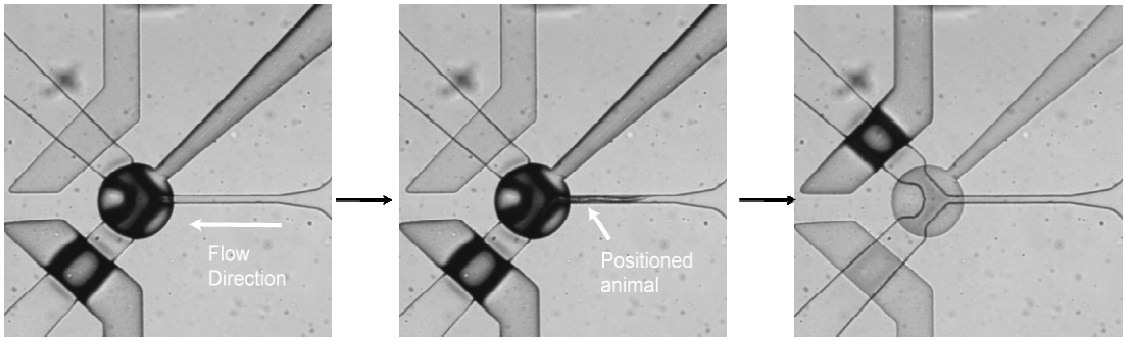



**Figure 17: Frames from videos showing device operation during screening. (left) valves set to allow a small amount of flow and wait for animal to enter. (middle) an animal that has been stopped in the imaging channel and reduces the flow by the increase in resistance. (right) the animal being sorted out of the bottom channel.**

Although sufficient for a screen performed by a human user that is capable of performing relatively complex error handling and adapting to changing conditions with ease, this device would be insufficient for automated screening. The extra complexity of an additional valve or two is more than warranted if it serves to decrease the error rate and allows continuous processing. Especially when performing rare sorting events, as would be the case in a forward genetic screen where 1 in 100-500 animals may be of interest, it is critical to minimize the sorting error. Future devices would require the addition of a mechanism to ensure loading of one and only one worm.

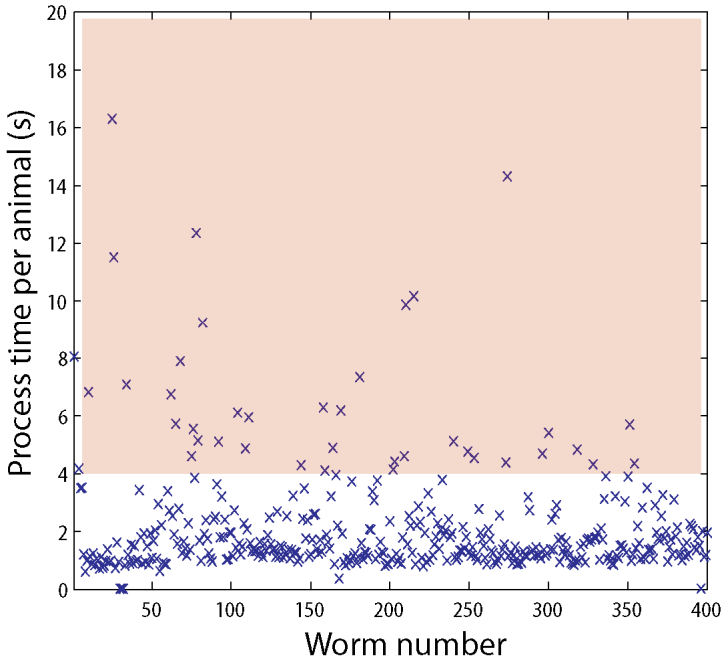



**Figure 18: A representative sequence of total processing time per animal, showing robust and easy animal handling and processing in the device. Animals of potentially interesting phenotypes are examined in detail, typically taking more than 4 seconds each (shaded in pink), while the majority of animals are processed in <2 seconds.**

## IMMOBILIZATION

This device has no method of immobilizing animals, other than restricting the animal motion within a channel on the same order as the worm itself. This restricts the motion of the worm, and when combined with fluid flow pushing the animal into the partially closed valve, provides a method to keep the worm within the field of view. This, however, is only sufficient for performing screens at medium resolution and numerical aperture. At low or medium resolution, human decision making is relatively unaffected by small amounts of motion. Attempts at using this system for performing screens with objectives with numerical apertures >1, failed, however, as the depth of focus was too small, and the animal motion obscured any of the interesting features. Screens based on dim fluorescent reporters would require a robust method of immobilization.

## SORTING

This device experienced some of the same problems in sorting as did the first generation device by Kwanghun Chung. In particular, there was occasionally a difficulty of sorting larger worms in a consistent manner. Because all screens with this device were performed at a relatively low magnification so the exiting channels could be seen, if an animal failed to exit properly, it could be easily seen and corrected by the user. For automated screening, however, this too would have to be corrected to ensure that all animals exited rapidly and without difficulty.

## CHARACTERIZATION OF NOVEL MUTANTS AND PERFORMANCE

During the screening 11 animals with unique synaptic expression were identified. These animals tended to have increased numbers of puncta along the nerve cord, and a decreased level of expression in the cell-body. Due to the amount of time required to perform the genetics, and identify each of the genes affected during the screen, the animals will be investigated over the next several years.

This screen helped to further understanding of the role of Rab-5 in neurons. In collaboration with Dr. Sharon Sann, a postdoctoral student at UCSD, one mutant discovered in the screen has been mapped and cloned. The mutation was found to affect the *C. elegans* Rabx-5, a guanine exchange factor (GEF) of RAB-5. *Rabx-5* mutants exhibit moderate aldicarb resistance, indicating defects in synaptic transmission. Quantitative imaging analysis of YFP::RAB-5 reveals specific alterations in synapses and cell bodies. In contrast, *rabx-5* mutations have little effect on the expression of constitutively active Rab-5-GTP (YFP::Rab-5(Q78L)). These results support that RABX-5 is a functional GEF for Rab-5 in neurons. Fluorescence recovery after photobleaching analysis suggests that RABX-5 could have differential activity in synaptic and extrasynaptic regions. Examining expression of the synaptic vesicle markers synaptobrevin

SNB-1::GFP and mCherry::Rab-3 reveals increased expression in synaptic regions in *rabx-5* mutants. Dr. Sann and Dr. Jin are currently examining the subcellular localization of RABX-5 and defining the regulatory mechanisms involving additional regulators of synaptic endosomes.

The computer-enhanced microfluidic approach demonstrated here has many advantages: 1) computer-assisted screening to accentuate phentoypical characteristics, which may be missed by manual screens, 2) human decision-making to allow flexibility if presented with a novel phenotype, 3) preconfigured image processing modules for minimal algorithm-development time, 4) at least an order of magnitude greater throughput than current manual screening, 5) higher magnification, higher numerical aperture optics than commercial or some manual screening systems, 6) almost three orders of magnitude less expensive than commercial systems, and 8) simple assembly and operation for use by technicians with little or no familiarity with microfluidics. These advantages should enable new types of screens in the near future.

## INCORPORATING FEATURES FOR AN AUTOMATED SCREENING DEVICE

In order to create a device that was capable of performing automated screening, a third-generation device was designed. The objective of this device was to maintain the same focus on simplicity that allowed the second generation to work so well, while adding features that allowed robust immobilization, consistent loading and sorting. The completed device is shown in Figure 19, and a zoomed in view of the active region of the device is shown in Figure 20. In this image the flow is from right to left.

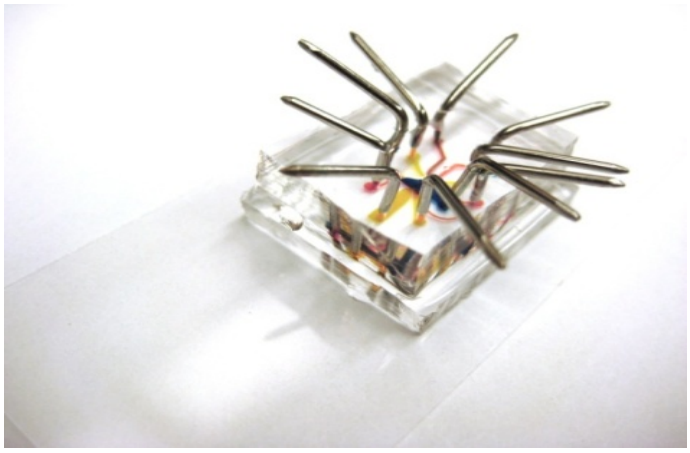


**Figure 19: A photograph of the dye filled device. Pins have been inserted into the appropriate fluid entrances.**

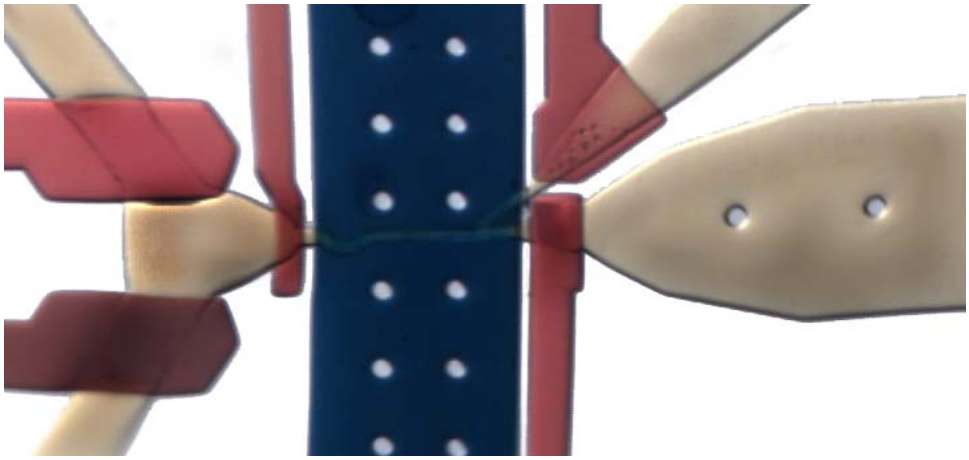


**Figure 20: A dye image of the third-generation device. This device was ultimately used to perform large-scale automated forward screens. Valves are filled with red dye, the cooling channel is filled with blue, and the yellow is the fluid flow channel for the worms.**

## OPTIMIZING THE IMMOBILIZATION THROUGH COOLING

As was previously discussed, one of the primary challenges with the first generation of devices, was inconsistent cooling across the length of the worm because of the dimensions of the coolant channel. By removing the suction channels, the dimensions of the cooling channel were increased so that it covered the entire worm. Additionally, this design of the original design featured several

## USING A FLUSHING CHANNEL TO ENSURE CONSISTENT SORTING OF ANIMALS REGARDLESS OF SIZE VARIATIONS

The first and second generation devices both experienced similar problems with the animal sorting. The problems were:

1) Changes in the time required for an animal to exit because of the increase in animal size. Once an animal touches the walls of the imaging channel, there is an exponential increase in the fluidic resistance, and thus the amount of force (and time) required to push an animal out of the imaging channel.
2) Changes in the flow rate through the entire device caused by resistance changes elsewhere in the device. Because the device relies on pressure driven flow, a change in resistance anywhere in the device affects the flow rate through the entire system. In ideal conditions, the resistance of the imaging channel dominates, but in reality this might not be the case.

Fluid resistance and flow is often modeled using the electrical comparison to voltage, current and resistance where current is flow, voltage is pressure and resistance is fluid resistance. A model of the original device in Figure 21 reveals the problems that occur during these two scenarios. In an ideal scenario, the worms are roughly on the same size scale, and the resistance of the imagining channel and the presence/absence of a worm dominates the resistance of the whole device. In this situation, a larger

or smaller animal may change the resistance and flow rate by one- or two-fold, but this largely doesn't change the device behavior. During typical device operation, the density of animals upstream of the imaging channel tends to fluctuate. As the density increases, the fluidic resistance increases, and the flow rate through the device drops. When the resistance upstream becomes greater the resistance of the imaging channel, the flow drops significantly and the device behavior becomes unpredictable.

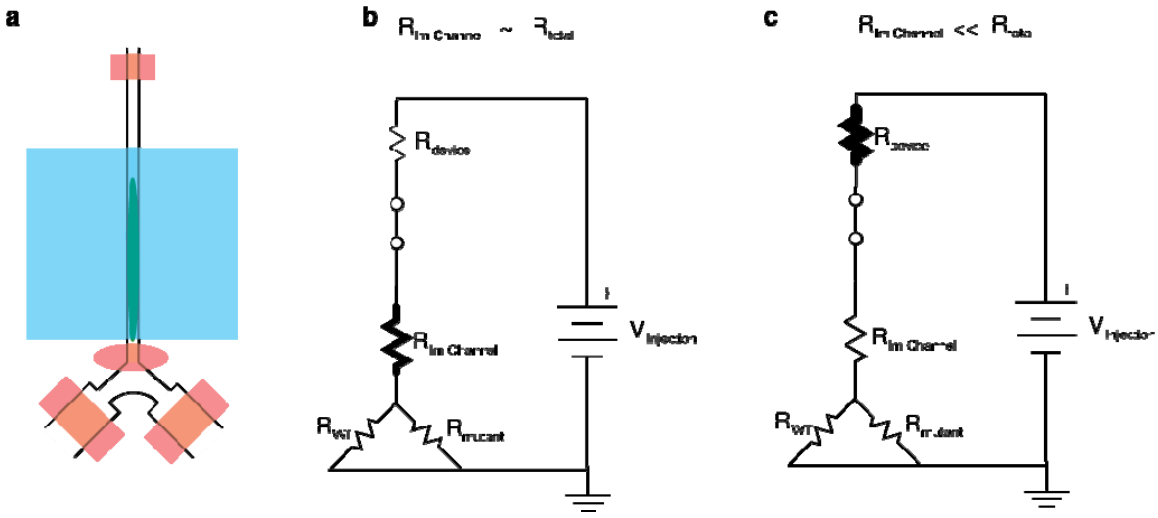



**Figure 21: A diagram of the first generation device and electrical analysis of the flow. A) A cartoon schematic showing the imaging channel and the exit channels. B) The ideal operational scenario for the device. In this case, the small area that is open underneath the positioning valve creates a very large amount of resistance. During the animal loading situation, this resistance dominates and the fluidic resistance of the device is roughly equivalent to the resistance of the imaging channel. This results in a relatively consistent fluid flow rate regardless of the size of the worm. C) In a non-ideal situation, the resistance of the up-stream device is greater than the resistance of the imaging channel. This happens if the density of worms in the device significantly increases. In this case flow is unpredictable, and could be an order of magnitude or two less than in the ideal situation. This causes significant problems in the worm exiting.**

In order to ensure that each animal was released into the appropriate exit channel in a consistent amount of time regardless of the animal size, or the fluidic resistance of the other regions of device, a flushing system was created to remove animals. This uses a high pressure (typically twice the pressure of the injection pressure) to push the animals out of the imaging channel. Using both on-chip and off-chip valves, this high-pressure flow is kept isolated from the system until an animal needs to be released. At this point the loading regulator valve is closed and the high pressure valve opened. This forces the majority of the flow through the imaging channel, and rapidly pushes the imaged worm into the appropriate channel.

The diagram in Figure 22 shows the various steps of the loading, imaging and sorting process for the third generation device. By including a flushing channel and pressure source, the resistance changes in the upstream device are neglected and the

Loading: $Flow = \frac{V_{injection}}{R_{device}+R_{im\ channel}+R_{exit\ channel}} \approx \frac{V_{injection}}{R_{device}+R_{im\ channel}}$

Exiting: $Flow = \frac{V_{injection}}{R_{im\ channel}+R_{exit\ channel}} \approx \frac{V_{injection}}{R_{im\ channel}}$

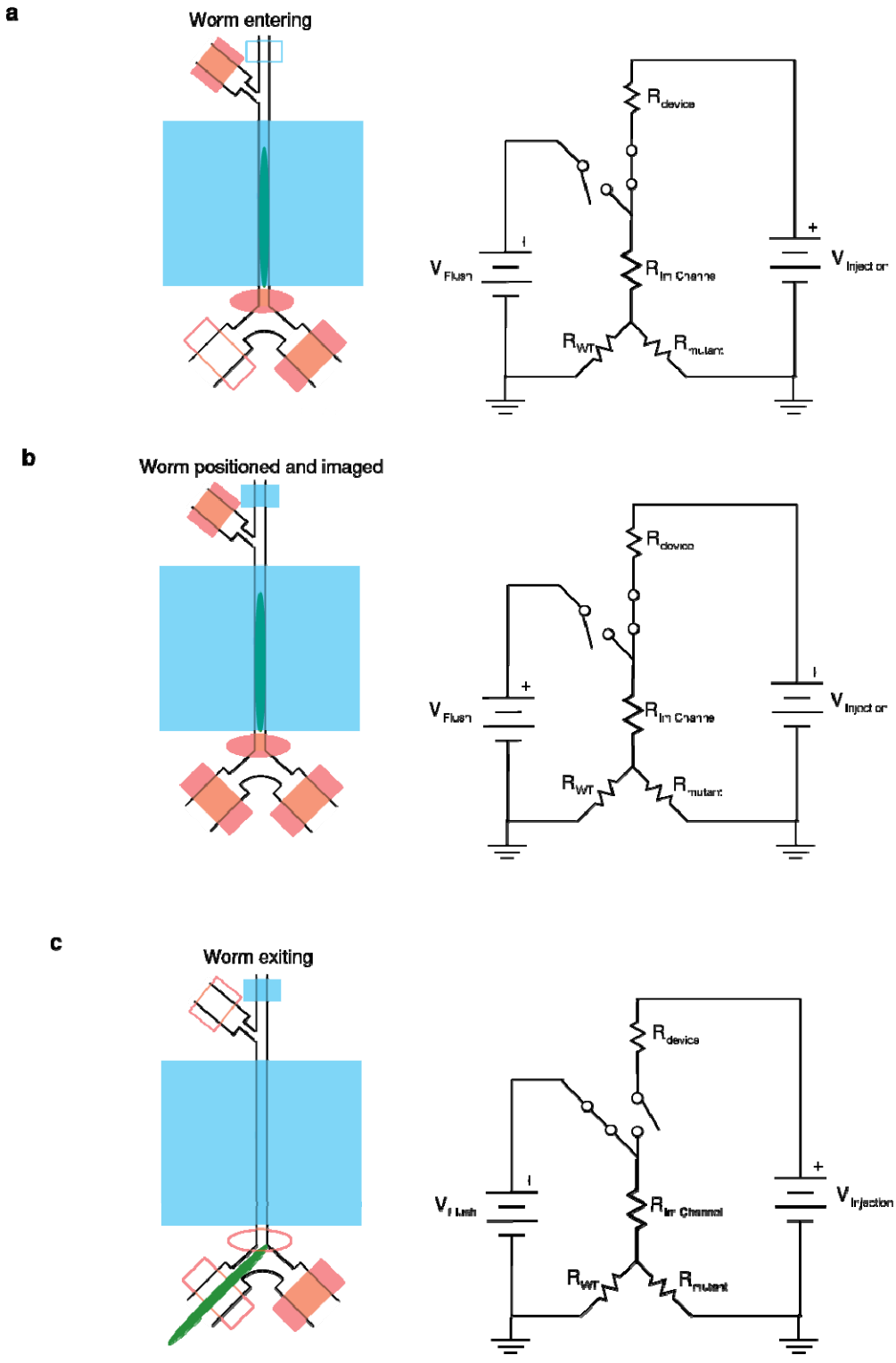



Figure 22: Schematic showing a cartoon of the device operation (left panels) and the electrical model for the system at each step.

## AUTOMATED DEVICE PERFORMANCE

The original device created by Kwanghun Chung had several problems that limited fabrication yield and prevented automation. These can be seen in Table 1, and are addressed at length in Chapter 1. The modifications that were made to address these weaknesses are listed in Table 2. Of particular note is that the mean time between device failure increased >450% as a result of the changes. Also, the addition of the flushing channel and fluid source not only increased the consistency of animal exiting, but allowed error handling routines to bet created. This combination increased the amount of time that the system could operate without human intervention by 1800%.

**Table 2: A summary of the problems with the first generation microfluidic device and the modifications that were made to resolve them. The resulting device had significantly higher yield rates and could be operated for extended periods of time when compared to the original device.**

| | Solution | Results |
|---|---|---|
| Difficulty of fabrication (low fabrication yield) | Modify design to allow increased rotational/translational alignment. Use partial closure valves to allow thermal bonding. | 50% yield increased to >90% yield |
| Suction channels design | Employ partial closure valves to increase the size of the smallest feature and reduce the likelihood of clogging. | Meant time between device failure increased from 1.5 hours to 7 hours |
| Inconsistent device operation with varied worm density | Use alternative pressure source for animal exiting and isolate from the rest of the device. | Allows error handling and increased the mean time between operator intervention from 15 mins to 4.5 hours. |
| Loading regulator works only in extremely idealized situations | Increase valve dimensions and use to isolate chamber during animal exiting | Allows error handling and increased the mean time between operator intervention from 15 mins to 4.5 hours. |
| Temperature gradient across animal | Remove suction channels to allow even coolant flow across entire animal | Temperature gradient across animal reduced from 7°C to 1°C |

# CHAPTER 3: CREATING A SYSTEM FOR AUTOMATED SORTING

In addition to the development of microfluidic devices, considerable engineering and design was required to create the external systemic components to allow for automated sorting. This required a closed loop control system and the development of specific hardware to interface with and control the on-chip components. Because microfluidics is a relatively nascent field, and currently confined primarily to academic research labs, there is a dearth of commercially available components to support microfluidic work. Practically, this has resulted in each lab creating customized tools to control their own devices. Creating external components that would allow computerized control of on-chip components was necessary. The requirements for these systems were similar to the microfluidic device of the previous chapter: robustness and allowing extended operation. The creation of the computerized control components was performed in collaboration with Jeffrey Stirman, and to a lesser extent, Kwanghun Chung.

## COMPUTER CONTROLLED VALVE ACTUATION

Controlling the pressure flows to the micro-valves on-chip was a critical element of the systemic development. Creating a robust system that both allowed pressure regulation and actuation of the pressure allowed complex valve sequences, and valves to be actuated within tens of milliseconds for hours on end. This system began as a relatively bulky arrangement of valves, regulators and fluid lines that took up a large portion of the optical table holding the microscope; however, it was eventually reduced in size and packaged as a single box. This allowed for duplication and distribution to collaborators and has increased the impact of the work by allowing adoption of these methods in non-microfluidics laboratories.

## PRESSURE REGULATION, ACTUATION AND CONTROL

To control the microfluidic system, a minimum of two regulated pressure sources is required, one for the valves and one for the animal injection. In many cases, however, three or four regulated pressure sources are required. The third-generation device would require a pressure source for injection, one for the flushing fluid and a third for the valves. Depending on the worm size, and the individual device batch, it is often necessary to use different pressures for injection or the valves.

Actuation of the pressure lines to the on-chip valves was done using miniature solenoid valves. The valves selected for the original system were based on recommendation from the Quake lab and were from the Lee Company. Further investigation led to a comparison between several manufacturers of small solenoid valves including Hargraves and Asco. Ultimately, the Asco Series 188 valves (Part #18800056) were selected for inclusion in the packaged system as they were significantly more cost competitive, they withstand pressures up to 115 psi (double the competitor valves) and includes a well designed pressure manifold.

Computerized control of the valves was done using a DAQ unit from national instruments. Because of the relatively low actuation rates required by the system, at most a few hundred Hz, a USB unit was selected. Inexpensive data acquisition units such as the NI 6501 we selected, can only provide a few mA of current per line, or less than a tenth of the amount required to actuate even the low power miniature solenoids. A small PCB to act as a voltage buffer for each of the lines was created.

## PACKAGED SYSTEM

The control system went through several evolutions before it was finally packaged. The packaged system (Figure 23) was designed a metal enclosure from McMaster, and provides enough flexibility to be used with any number of microfluidic systems. The finalized specifications for the system are:

- (2) 0-60 psi pressure regulators and indicators to control the on-chip valves
- (8) computer controlled solenoid valves for on-chip valve control. 4 valves for each of the 2 pressure regulators mentioned above.
- (2) 0-15 psi pressure regulators and indicators
- (2) computer controlled solenoid valves for injection. These valves are set to the same pressure
- (1) computer controlled solenoid valve for the flushing pressure output
- (4) powered outputs that are computer controlled for external solenoid valves in case of future expansion

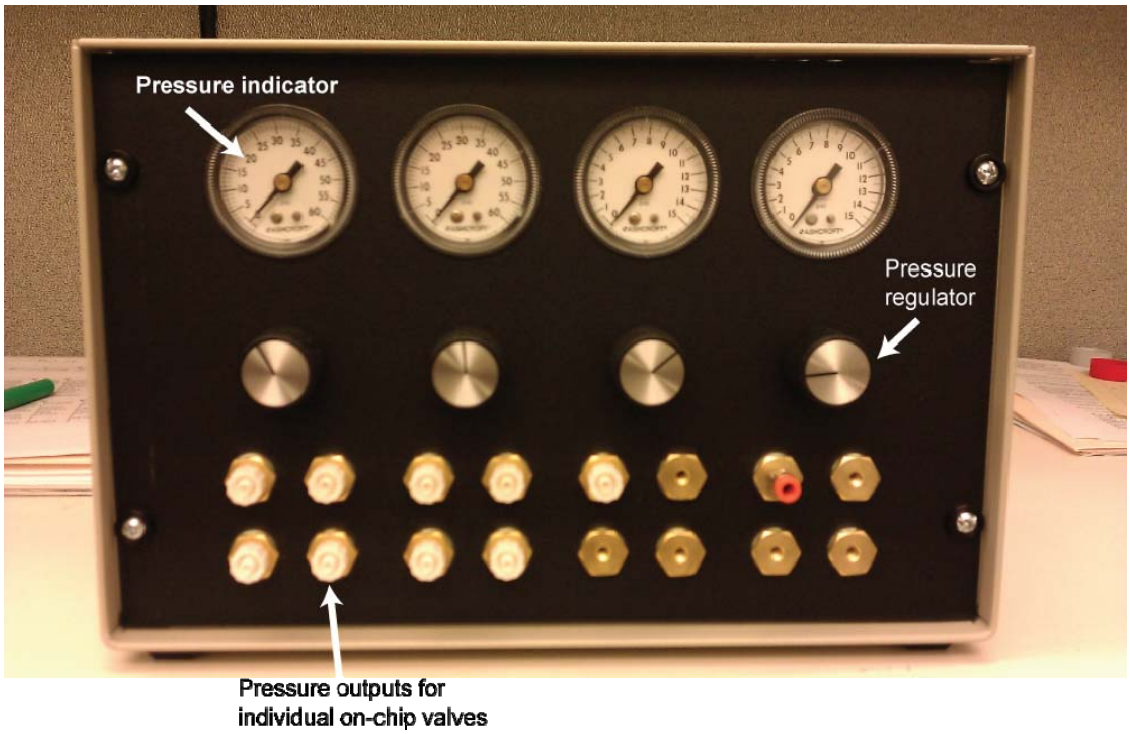



**Figure 23: The exterior of the completed control box. The box itself is made out of machined sheet metal, and contains four pressure regulators and indicators. The left two regulators are for use with the on-chip valves, and the right two are for use with the animal injection and flushing systems.**

The inside of the control box is shown in Figure 24. Rather than using barbed tube connectors, as was originally done, all of the gas connections are done with push-to-connect fittings. This dramatically reduced the failure rate, and made assembly and diagnosing problems much simpler. Because of the

relatively slow response time of the on-chip valves (tens or hundreds of Hz), a USB DAQ board was found sufficient. To date, two of these boxes have been created and setup in labs with whom we are collaborating. Until microfluidics gains a significant commercial foothold, and systems like this are available for purchase, this is the only option of most biology labs interested in using the tools we have developed. This has magnified the real-world impact of the work in this thesis several fold.

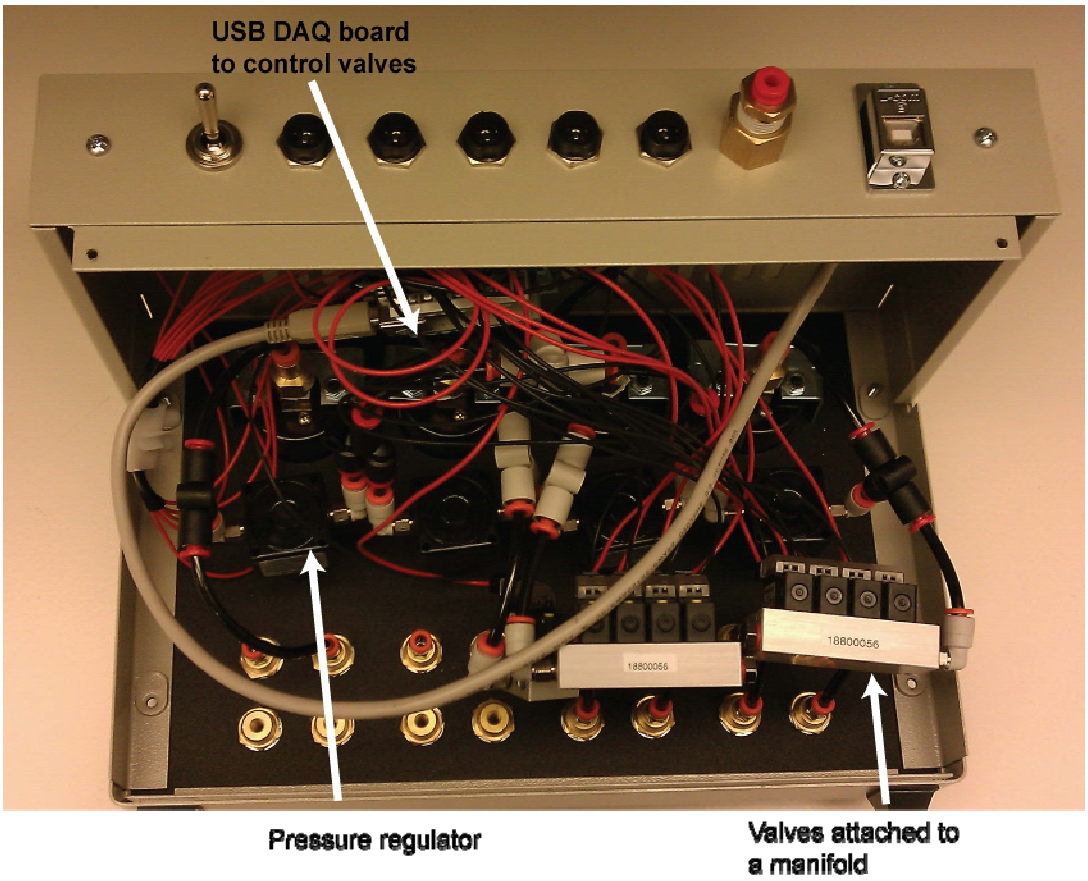



**Figure 24: The internal components of the control box. Inside the box are the pressure regulators, 12 computer controlled solenoid valves, and a USB DAQ board for computer control. The box requires a single 12V power source, a single pressure source of 60-100 psi, and a single USB line.**

## COMPONENTS FOR INJECTING ANIMAL SUSPENSION

Animal handling is the second critical component of the system development. The animal handling portion of the system can be separated into two portions: injection of a mixed solution of animals and

buffer, and a system to inject buffer to flush animals out post screening. Although this required two distinct fluid injection systems, they utilize the same underlying principles and are thus grouped together. Both systems are designed around a small, pressurized fluid container that contains tubing at the bottom of the container. When pressurized, the fluid is driven into the tube and thus into the microfluidic device. A schematic demonstrating this is shown in Figure 25.

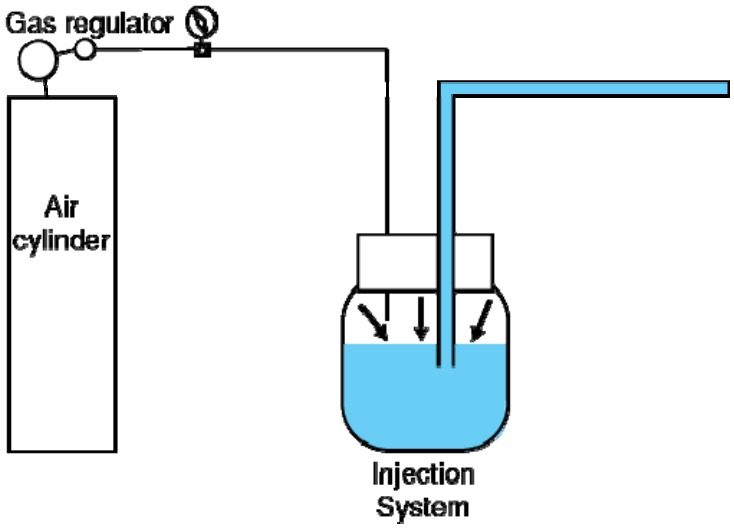



**Figure 25: A schematic of the injection system. A high pressure gas is used to pressurize the container. This in turn drives the fluid out of the container.**

## IMMOBILIZATION

The microfluidic system requires a method of immobilization that would be capable of completely immobilizing the animal for several seconds in a reliable manner. Rather than relying on the conventional method of anesthetics, which can potentially affect synapse expression and morphology, we chose to cool the animals. Cooling the animals to a low temperature (<5°C) was found to immobilize the animals in a manner that was completely reversible, and as long as it is short term, appears to have no effects on

normal *C. elegans* lifespan or behavior. Most importantly, using cooling to immobilize animals was found to have no effects on the morphology of the fluorescent reporters(66).

## COOLING SYSTEM

The unique nature of the microfluidic system meant that a specialized cooling system had to be designed that would capable of providing a coolant to the device with a consistent cooling capacity. The cooling system had to meet the following criteria:

1) Able to cool small amounts of fluid to ~-10°C.
2) Possess a relatively small physical footprint so that it could be placed on an optical table
3) Provide a relatively low fluid flow rate (typical flow rate ~1 ml/min)

These specifications differ from anything that is commercially available, and resulted in an extensive period of research development to create a system that met these objectives.

Because the rest of the fluid flow is pressure driven, all early implementations used a similar pressure driven system. This was a source of significant error, however, as the temperature of the coolant within the device and the amount of heat the coolant can remove from the device are based largely on the flow rate. The small thermal mass of the coolant, however, means that it approaches room temperature extremely rapidly once it has left the heat exchanger. This wouldn't be problematic if it was possible to measure the temperature of the coolant on-chip, but this isn't feasible in a reasonable manner. Any on-chip measurement would significantly alter the device design. Additionally, it would increase the cost of each device by several orders of magnitude, and the fabrication time required.

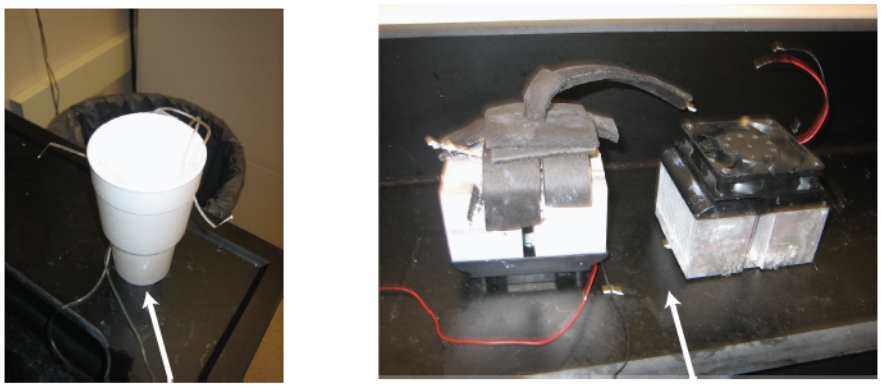


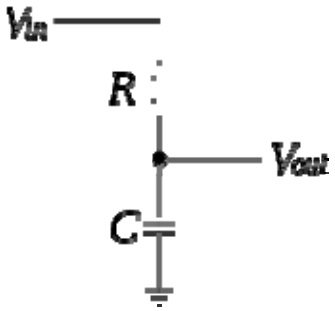

**Figure 26: Early cooling attempts. (left) a length of tubing was immersed in an ethanol and dry ice bath. Although the ethanol and dry ice mixture remained at a constant temperature, it proved unwieldly and inconsitent. (right) a two stage system using a system similar to the final method.**

Early coolant methods (Figure 26) focused the development of everything from circulating ethanol cooled with dry ice, to a simple ice water bath. Much of the preliminary work focused on using copper tubing pressed against the Peltier cooler as the heat exchanger. The coiled copper tubing was unable to efficiently used the surface area of the Peltier cooler and provide a sufficient dwell time for the fluid. This led to a downward spiral of applying increased voltages to the Peltier cooler, which placed increased thermal dissipation requirements on the heatsink. Unable to keep the "hot" side of the Peltier near room temperature, these systems proved unable to provide the requisite cooling. Ultimately the system was created using the following design principles:

1) A coolant solution using a solution of water - and either ethylene glycol or glycerol. In contrast to many other fluids (i.e. ethanol), neither of these diffuse through PDMS(129).
2) A peristaltic pump to provide a constant flow rate through the device. This ensures that regardless of resistance changes in the device or the viscosity changes in the fluid, the fluid will be exposed to the heat exchanger for an equal amount of time, and will take an equal amount of time to reach the device. Early attempts relying on pressure driven flow lacked this ability.

Furthermore, a peristaltic pump allows recirculation of the coolant allowing for system operation for hours at a time.

3) A capacitance device. One of the problems with peristaltic pumps is the pulsatile fluid flow that results. In practice this resulted in a ~1.2 Hz vibration of the imaging channel and made imaging impossible. A device with a large amount of capacitance was created, and placed in-line. This acts as a low pass filter and removes any vibrations in the fluid flow. The low pass design can be seen in the schematic below.




4) A custom designed heat exchanger attached to a Peltier cooler. Made out of machined copper for a high thermal conductivity, and designed so that the fluid had a surface area to volume ratio to allow for rapid heat exchange.

5) A temperature controller from Omega (CNI3252-DC), was used to ensure a constant temperature of the heat exchanger, regardless of the room temperature. A thermocouple placed in the heat exchanger is used, and the voltage supplied to the Peltier cooler is correspondingly increased or decreased.

These components can be seen in the diagram of the system below (Figure 27).

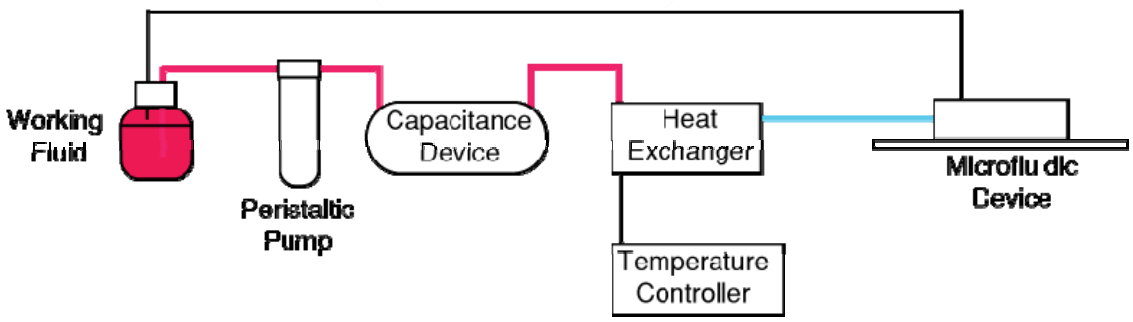



**Figure 27: Schematic of the finalized cooling system. By using a closed loop temperature controller, the system maintains the heat exchanger at a constant temperature and ensures that the system is consistently cooled. The peristaltic pump, coupled with the capacitance device, maintains a consistent, smooth, flow rate.**

Commercially available systems to cool small amounts of fluid to the required temperature, were not available. Because of this, the heat exchanger, and the cooling components had to created independently and with custom specifications. A thermoelectric (Peltier) cooling element was selected based on the small footprint and high temperature differential that could be created. Peltier coolers work by creating a temperature differential from one side of the cooler to the other, and thus it is necessary to dissipate a significant amount of heat in a small space. The temperature differential is dependent on the amount of heat transferred from cold to hot and the applied to cooler. During typical operation, ~15W of were dissipated on the “hot” side of the cooler. The copper heat exchanger maximized the use of the Peltier surface area and increased the dwell time of the fluid significantly.

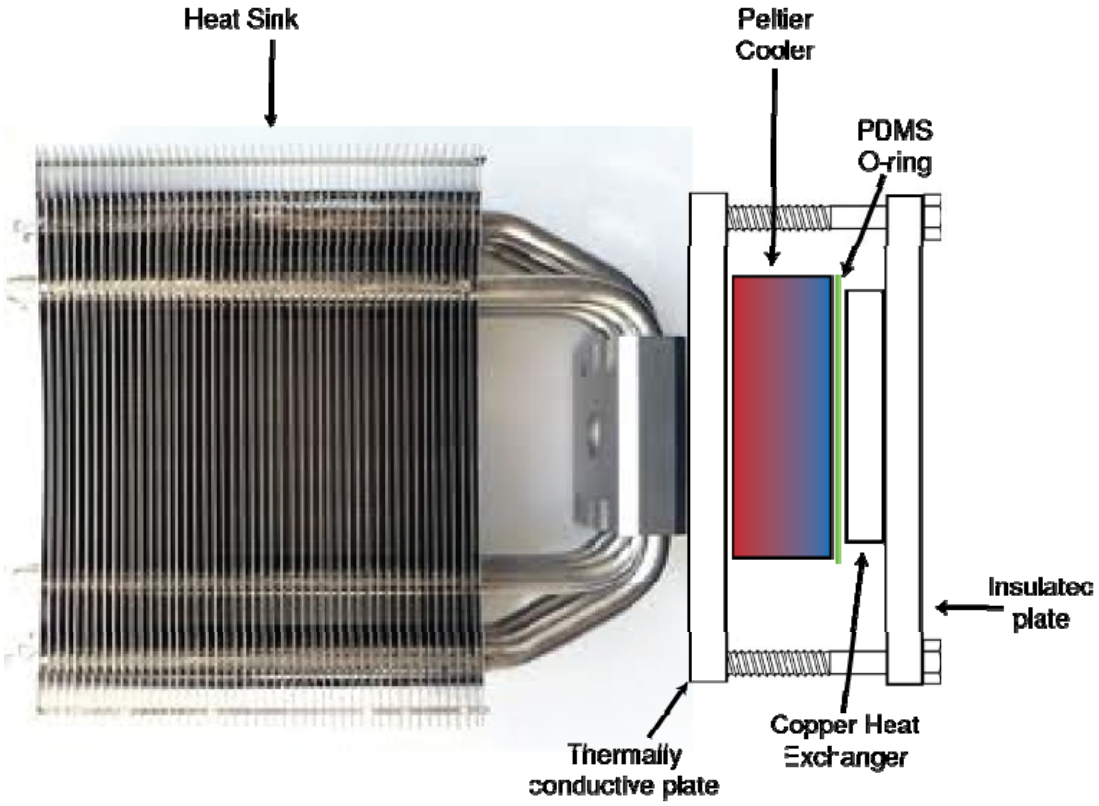



**Figure 28: A schematic of the completed heat exchange system. A custom machined copper heat exchanger is clamped onto a Peltier cooler. Before clamping onto the Peltier cooler the surface of the heat exchanger was covered in a thin layer of PDMS to act as an O-ring and prevent leakage. The coolant flows directly on the surface of the Peltier cooler. A commercial computer heat-sink is used to remove the excess heat from the Peltier cooler.**

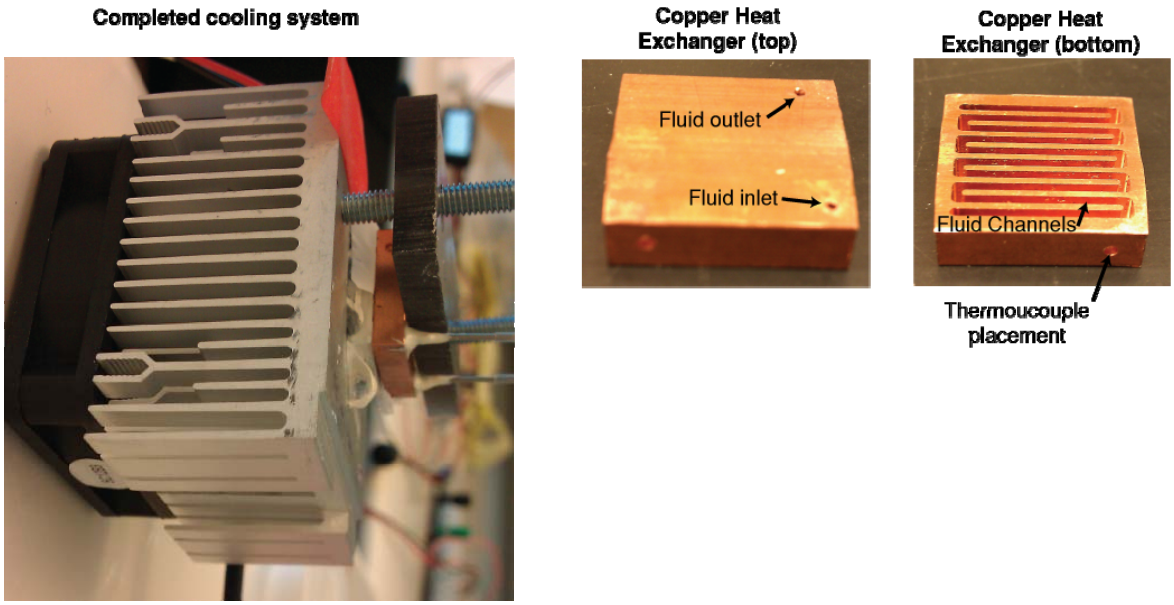



**Figure 29: Pictures of the completed cooling system. (left) the completed cooling system with the heat sink, Peltier cooler and copper heat exchanger. (right) The custom designed and machined copper heat exchanger.**

## COOLING EFFICACY AND BIOLOGICAL CONSEQUENCES

To validate the ability of the cooling to immobilize animals, and to compare it to an alternative method, I acquired video of an animal with a synaptically localized fluorescent reporter. After ten seconds of cooling, the animal hasn't moved at all, but in the alternative immobilization method you can see dramatic shifts in the synapses and fat granules.

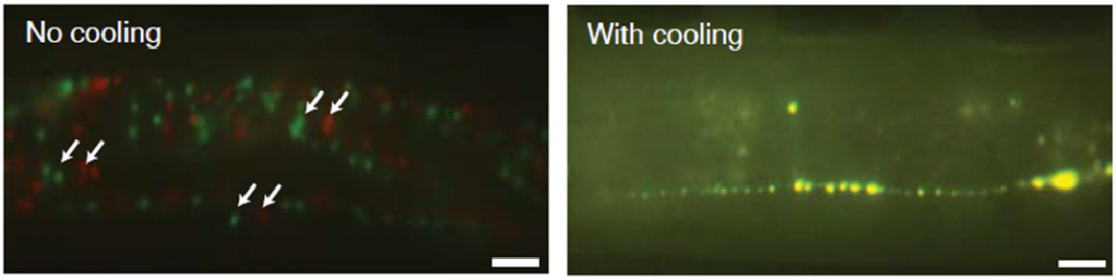



**Figure 30: Overlaid fluorescent images of a mutant animal taken at high magnification demonstrating the necessity of cooling: (Top) Two frames taken of a worm mechanically clamped but with no cooling. The first frame was colored red and the second frame 270 msec later was colored green. Sets of arrows demonstrate that same features moved significantly in the 270 msec between the two frames. (Bottom) Two frames 10 sec apart (also colored red and green) of a worm imaged with cooling, showing no discernible movement as the image is largely yellow.**

Verification that the short-term cooling had no effect on the animals was done in two ways. The first was to video-record animals exposed to short term cooling and phenotype their locomotion. Comparisons between wild-type animals that were processed through the device and cooled for 5 seconds each, to wild-type animals not cooled in the device, showed no discernible behavioral differences. Secondly, synaptic patterns of a fluorescent reporter were compared between animals immobilized with a common anesthetic (sodium azide) and animals immobilized and imaged in the device. As can be seen in Figure 31 there were no discernible differences between the synaptic patterns of animals depending on the immobilization method used.

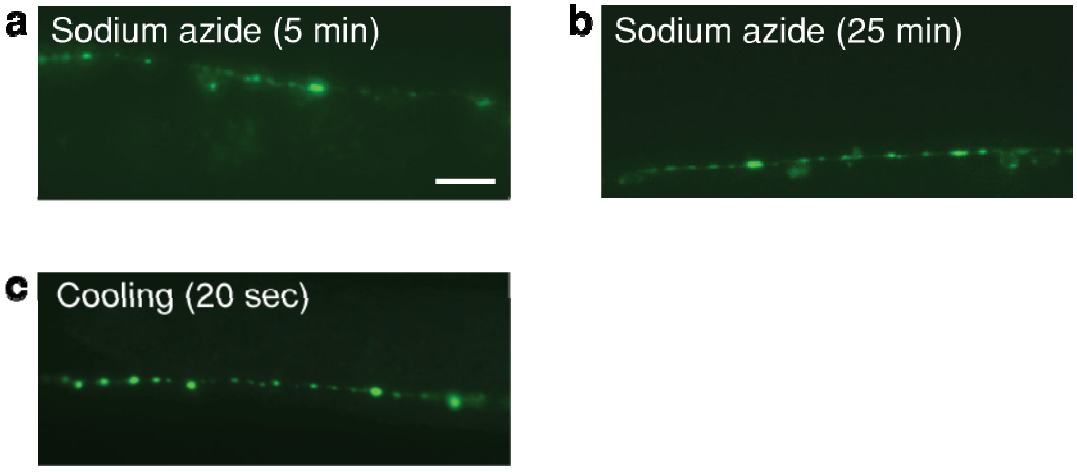



**Figure 31: Representative images of YFP distribution of an animal containing the transgene juIs198 [punc-25-YFP::rab-5] and a background mutation in unc-16(ju146). Animals immobilized with either (a-b) 10 mM of sodium azide or (c) cooling, showing similar punctal patterns. Scale bar: 10 µm.**

## COMPUTERIZED CONTROL SUFFICIENT FOR SORTING KNOWN, RELATIVELY SIMPLE PHENOTYPES

To demonstrate the effectiveness of the systems-level components, and begin development of a computerized system for automated screening, an intermediate framework for automated sorting of simple phenotypes was developed. I developed a software system comprised of control, acquisition, and processing modules, and applied it to sorting several different phenotypical expressions at different

magnifications and with differing complexities. This was the first demonstration of using an automated system to perform imaging and sorting of a multicellular model organism. The phenotypes were selected to allow relatively easy decision making for simple sorting. This allowed development of code processing animals, and allowed us to test the system operation under different scenarios to determine the sorting accuracy in relatively simple sorting situations. The results from the three separate sorting experiments are shown in Table 3, below.

**Table 3: Results of the high-throughput automated sorting experiments based on cellular and subcellular features, showing the number of each type of animals in the input and outputs. The rates of false positive, false negative and enrichment are calculated from these numbers. PQR-on vs. PQR-off: sorting of PQR GFP-positive animals versus PQR GFP-negative animals, demonstrating efficient sorting and low false negative rates. WT (1-AWC-on) vs. slo-1 (2-AWC-on): sorting based on number of AWC-on neurons, demonstrating the enrichment of mutants and the very low false negative rates. WT vs. unc-16: sorting based on synaptic features, showing efficient sorting and low false negative rates.**

| | Trial | Total number input | | Sort output 1 | | Sort output 2 | | Total number sorted | Percent in mix | False positives (%) | False negatives (%) | Enrichment (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $GFP^+$ as $GFP^+$ | $GFP^-$ as $GFP^+$ | $GFP^-$ as $GFP^-$ | $GFP^+$ as $GFP^-$ | | $GFP^+$ | | | |
| PQR $GFP^+$ | 1 | ~2,500 | | 640 | 86 | 1,601 | 119 | 2,446 | 31.0 | 11.8 | 6.9 | 284 |
| versus | 2 | ~2,100 | | 629 | 62 | 1,314 | 79 | 2,084 | 34.0 | 9.0 | 5.7 | 268 |
| PQR $GFP^-$ [a] | 3 | ~2,500 | | 794 | 61 | 1,541 | 104 | 2,500 | 35.9 | 7.1 | 6.3 | 259 |
| | 4 | ~1,800 | | 508 | 39 | 1,176 | 73 | 1,796 | 32.3 | 7.1 | 5.8 | 287 |
| | Average | | | | | | | | | 8.8 ± 2.2 | 6.2 ± 0.6 | 274 ± 14 |
| | | | Percent mutant | 2-AWC-on as 2-AWC-on | 1-AWC-on as 2-AWC-on | 1-AWC-on as 1- AWC-on | 2-AWC-on as 1-AWC-on | | 2-AWC-on | | | |
| 1-AWC-on | 1 | ~1,200 | ~1.5 | 19 | 90 | 1,021 | 2 | 1,132 | 1.9 | 82.6 | 0.2 | 940 |
| versus | 2 | ~1,400 | ~1.5 | 16 | 110 | 1,277 | 1 | 1,404 | 1.2 | 87.3 | 0.1 | 1,049 |
| 2-AWC-on [b] | 3 | ~1,200 | ~1.5 | 14 | 77 | 1,091 | 1 | 1,183 | 1.3 | 84.6 | 0.1 | 1,213 |
| | Average | | | | | | | | | 84.8 ± 2.4 | 0.1 ± 0.1 | 1,070 ± 140 |
| | | | Percent mutant | *unc-16*$^{-/-}$ as *unc-16*$^{-/-}$ | Wild type as *unc-16*$^{-/-}$ | Wild type as wild type | *unc-16*$^{-/-}$ as wild type | | *unc-16*$^{-/-}$ | | | |
| Wild type | 1 | ~1,400 | ~30 | 410 | 2 | 930 | 29 | 1,371 | 32.0 | 0.5 | 3.0 | 311 |
| versus | 2 | ~1,400 | ~25 | 308 | 32 | 971 | 21 | 1,332 | 24.7 | 9.4 | 2.1 | 367 |
| *unc-16*$^{-/-}$ [c] | 3 | ~1,400 | ~25 | 309 | 37 | 1,030 | 24 | 1,400 | 23.8 | 10.7 | 2.3 | 375 |
| | Average | | | | | | | | | 6.9 ± 5.6 | 2.5 ± 0.5 | 350 ± 35 |

## EXPRESSION PATTERN ANALYSIS ENABLED BY LARGE SCALE RAPID PROCESSING

Gene expression pattern analysis is a common technique in genomic studies(130-132), as well as mosaic and genetic analysis(133), and gross phenotyping(134). Typically one would be interested in the morphology, intensity, location, and timing of appearance of a (fluorescent) reporter. Performing global studies of gene expression makes it possible to determine the functions and interactions of genes. Here, we demonstrate the ability of our system for rapid gene expression profiling on a population of animals carrying a reporter transgene *kyIs342*(135). In these animals, green fluorescence protein (GFP) is expressed in sensory neurons AQR, URXL/R and PQR*(48)*. GFP is expressed in coelomocytes (coinjection marker *punc-122::gfp*) and background fluorescence from the intestine is also observed (Figure 32a).

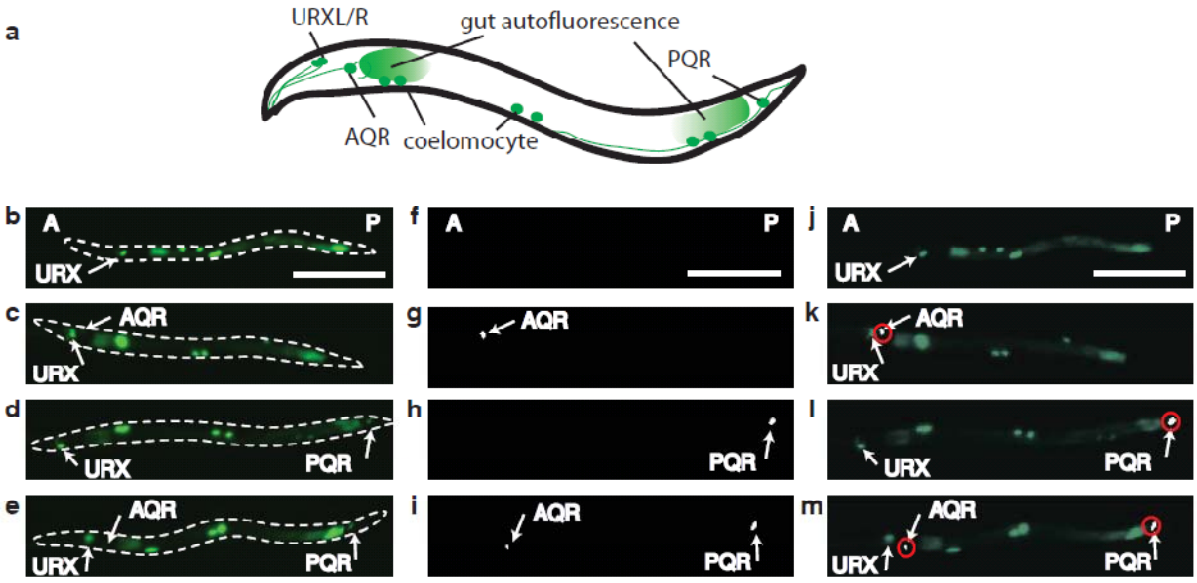



**Figure 32: Automated analysis of gene-expression pattern in the integrated micro system. (a) Schematic of the fluorescent regions including expressions of multiple GFP transgenes and autofluorescence from the gut. (b-e) Representative images showing stochastic expression of reporter genes kyIs342 [pgcy-32::tax-4::GFP, punc-122::GFP] in wild-type. (b) GFP in URXL/R only; (c) GFP in AQR and URXL/R; (d) GFP in PQR and URXL/R; (e) GFP in AQR, POR, and URXL/R. (f-i) Processed images showing the identified neurons, distinct from other fluorescence signals. (j-m) Overlay of the raw images (b-e) and the processed images (f-i). Scale bar: 100 µm.**

In this experiment, animals were injected into the microchip and the observation chamber one at a time (via the self-regulated loading mechanism). Animals were freely moving and no immobilization was used;

all images were acquired at room temperature, and at 10x. Our automated system acquired images of a large number of animals processed through the microchip continuously. The software performed image processing to distinguish specific neurons not only from background auto-fluorecence and coelomocytes, but from each other; additionally, the software also quantified the expression level in these neurons in a high-throughput manner. We found that expressions in the URX are consistent, but expressions in AQR and PQR are stochastic with four possible combinations: AQR+/PQR+, AQR+/PQR-, AQR-/PQR+, and AQR-/PQR- (Figure 32b-e). Further off-line quantification also shows that expressions of the reporter in AQR and PQR have different mean intensity as well as different distributions. Compared to manual microscopy, the distribution of the expression histogram is very similar and the throughput is significantly faster because both the loading of the animals and the image acquisition were rapid and wholly automated. During the experiment, the loading time into the observation chamber occurred within 1 second after the previous animal exited for over 58% of animals. This simple scheme allows us to process at a rate of >400 animals per hour continuously and automatically. In addition, because the animals remain at room temperature with no exposure to anesthetics, there should be no concern about the alteration of the expression pattern and it is possible to collect and reimage the animals at a later time. This mode of operation of our system can thus be adapted, for example, for the analysis of promoter activities, genetic interactions, and pathways in conjunction with reporters.

To analyze the images, out-of-focus frames were discarded and the images are convolved with *[1 1 -1]* matrix to accentuate small bright regions. A threshold was applied to determine the fluorescence from the intestine as well as AQR and PQR. Different thresholds are then applied to the left and right nematode centroid to identify AQR and PQR and to distinguish PQR from the intestine auto-fluorescence. Groups of remaining pixels are then compared based on a number of features (size, position, etc) to determine whether AQR and PQR are present, and if so, where they are located. For each animal, the most in-focus of the pictures is used to determine the intensity of AQR and PQR. The fluorescence

intensity could depend on the orientation of the immobilized animal, but does not significantly impact the quantification results for this and the following applications. The correctness of the output (the location and presence of AQR and PQR) for each animal was independently verified and corrected if necessary. The algorithm was found to have >95% accuracy.

## PHENOTYPING, SORTING, AND SCREEN BASED ON 3-D CELLULAR FEATURES

Forward and reverse genetics can result in alterations of genes and their expression levels to change detectable phenotypes in the animals. Full genome coverage usually requires examining large numbers of animals (~100,000) to find the rare few with dissimilar phenotypes. To demonstrate the microsystem's ability to perform an automated rare sort, we performed a mock screen by sorting a small number of mutant animals mixed in a background of animals of wild-type genotype. We used animals that express *str-2*::GFP in AWC neurons. This experiment also demonstrates the ability to screen using diffraction-limited optics based on signals from different cells at similar in-plane locations but different depth.

The anatomy of *C. elegans* has many bilaterally symmetric cells, which may however exhibit functional asymmetry. In wild-type animals, one of the two AWC neurons expresses *str-2* (AWC-ON). Mutations in *slo-1* gene produce the 2-AWC-ON phenotype(136), where both AWC cells express *str-2*. In our experiment, the microsystem sorts a small fraction of *slo-1* 2-ON animals from wild-type 1-ON-1-OFF animals based on z-stacks images (Figure 33). Because AWCL and AWCR are located at the same position along the anterior-posterior axis and are <20 microns apart, low magnification microscopy cannot distinguish 1-ON from 2-ON phenotypes. We therefore used a 100x oil objective (NA=1.4) and designed the microchip to be compatible with this type of high-resolution microscopy. Additionally, complete immobilization of animals is necessary because any movement would cause blur the images and make identification of the neurons impossible. This was accomplished by transient cooling of the animal to ~ 4 °C. Once the animal is immobilized, sparse z-stack images along the body of each animal are obtained to

determine the location of the head (Figure 33a). The stage then centers on what is identified as AWC neurons within the field of view, and a denser z-stack is acquired (Figure 33b). Further image analysis by flattening and thresholding is then used to determine whether the animal is 2-ON (*slo-1*, Figure 33c,d) or 1-ON-1-OFF (wild-type, Figure 33e inset) phenotype. The animals are then sorted according to their phenotypes. Quantifications of the marker intensity can also be performed (Figure 33e). The z-stack images of all animals are written to a hard-drive during the phenotyping and sorting to allow the user to retrieve them later for further analysis or manual verifications. During this experiment, most time was consumed by scanning the entire animal and saving images to disk; the throughput could be greatly increased by optimizing the procedure to locate the neurons and by reducing the number of images saved for each animal.

To locate the AWC neurons, sparse z-stacks (5 µm steps) are gathered along the A-P axis of the worm. It takes <10 seconds to image an entire L4 animal at 100x using a 5 µm step size. These z-stacks are then flattened by computing the standard deviation of pixel in the x-y plane along the z-direction. To ensure that intestine autofluorescence is not mistakenly identified as a neuron, intestine fluorescence is located and removed from the picture. A threshold is applied to this newly flattened image and the neuron(s) are located. The xyz stage then moves to the location of the neuron(s) to grab a more detailed z-stack (1 µm step size) which is necessary to determine whether one or two AWC-ON neurons are present. This z-stack is similarly flattened and a threshold is applied to determine the number of neurons. For the intensity histogram, we processed the images based on the average intensity of cell body from the brightest plane in a z-stack, and in effect the value is normalized against the cross-sectional area, which corrects for any rotational variations.

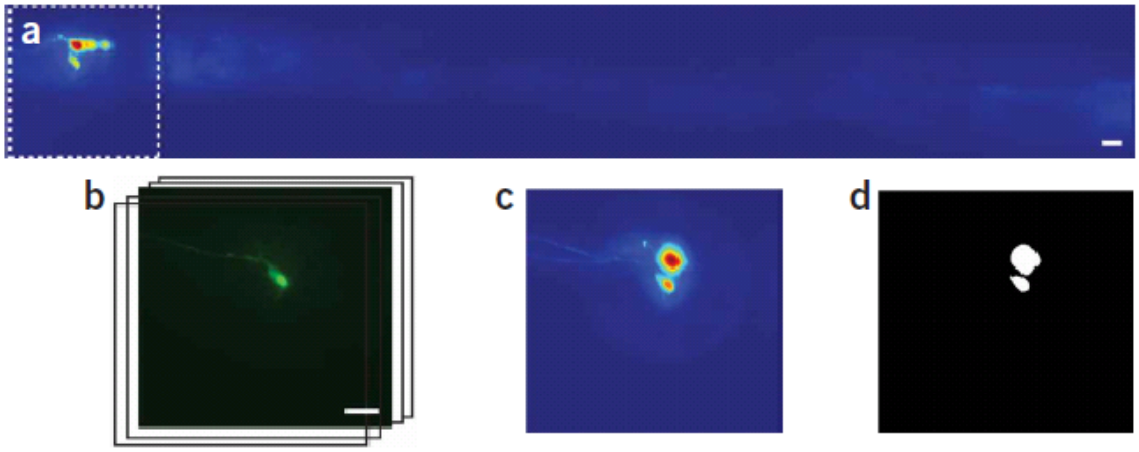



**Figure 33: Automated three-dimensional imaging and sorting with cellular resolution in the integrated micro system: image-processing and decision-making to sort animals based on the number of AWC neurons expressing pstr-2::gfp. This set of images corresponds to a slo-1 mutant with two AWC-ON neurons. (a) Flattened series of sparse z-stack images along the body of animals showing the cell bodies and the neurites. Red represents high contrast and blue represents low contrast. (b) Denser z-stacks near the head of the animals so that it is possible to establish the number of AWC-ON neurons. (c) Flattened dense z-stack images near the head. (d) Thresholded image showing identified neurons. Scale bars: 10 µm.**

Age-synchronized adult animals of ~1.5 % 2-ON *slo-1* mutants mixed in 98.5% wild-type (1-ON) background based on the GFP patterns (Figure 33f) were successfully sorted. With on-line processing and decision-making without human supervision, we achieved a throughput of ~one animal per minute. Additionally sorting accuracy was verified by examining the collected animals for roller phenotype, since the *slo-1* strain used in this experiment also carries *rol-6*. All but one 2-ON animals were sorted correctly; effectively we enriched the 2-ON animals by >25 fold. The false positive rate (1-ON-1-OFF animals being sorted as 2-ON animals) and the false negative rate (2-ON animals being sorted as 1-On-1-OFF animals) were 2% and 25% respectively with a total of ~270 animals input. If desired, accuracy of the sorting can be further improved by altering sorting criteria as well as implementing multiple rounds of sorting.

## PHENOTYPING, SORTING, AND SCREEN BASED ON GROSS SYNAPTIC FEATURES

As our understanding advances, genetic screens are becoming increasingly difficult to perform because the phenotypes of interest are becoming finer and subtler. For example, many of the synaptic or other subcellular reporters exhibit features that are sub-micron in size, dim, and easily photobleached. Because

of these characteristics, manual phenotyping is often not only too slow, but also subjective and cannot detect quantitative changes. Using our automated system, we eliminate the need to seek the targeted region in the sample, therefore greatly reducing the exposure time and the extent of photobleaching. By automated on-line image processing, we can also phenotype animals more accurately and at a greater speed. Additionally, synapse features that are highly sensitive to the application of anesthetics can be examined in our system by applying cooling instead.

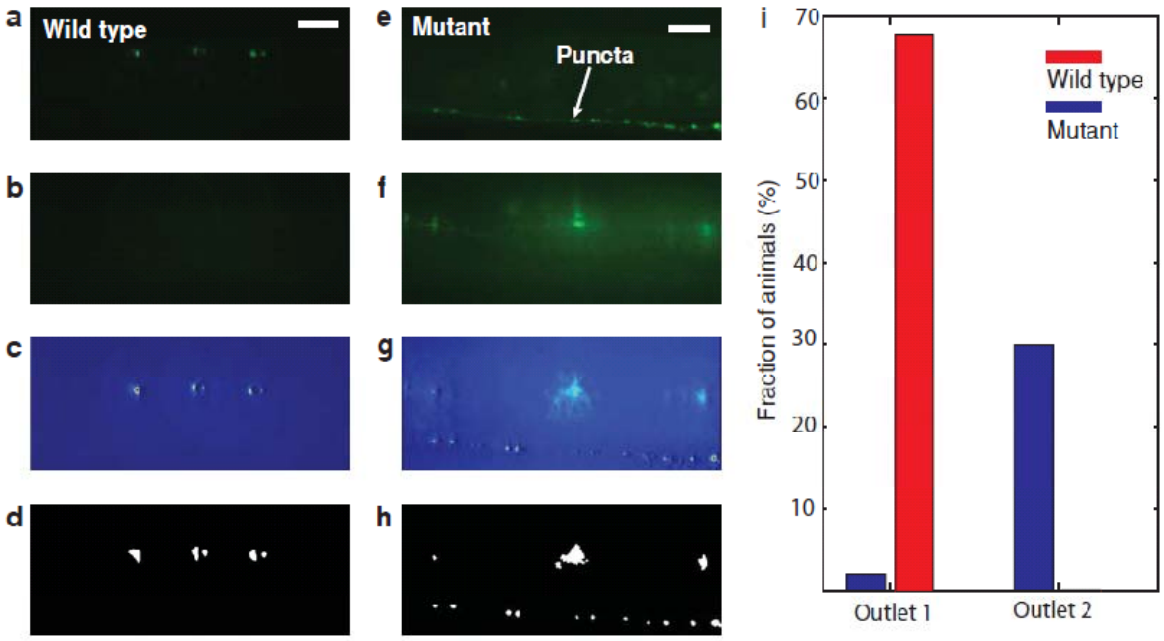



**Figure 34: Automated high-throughput microscopy and sorting based on synaptic marker phenotypes. (a, b, and e, f) Representative images of punc-25-YFP::RAB-5 reporter expression in animals at two focal planes 20 microns apart. (a, b) Wild-type. (e, f) unc-16 mutant. (c, g, and d, h) Processed images from the corresponding animals to find puncta along the nerve cord and cell bodies. Mutants have puncta structures along the nerve cord, which were the basis for sorting. (i) A graph showing the sorting accuracy.**

To demonstrate the ability to screen based on subcellular changes, we performed a sorting experiment using strains carrying an integrated reporter transgene juIs198 [punc-25-YFP::rab-5], which is expressed in the GABAergic motorneurons and labels a subset of synaptic endosomes(127). In the wild-type background, YFP-RAB-5 is faint in the nerve cord (Figure 34a,b). unc-16 mutants have an altered phenotype in the marker intensity along the nerve cord (Figure 34e,f) (Heather Brown and Yishi Jin,

personal communication). Age-synchronized mixed populations (~30 % mutants and ~70% animals with wild-type background, ~1370 animals total) were sorted according to their synaptic marker phenotypes at a rate of ~250 animals per hour. On-line image processing (samples shown in Figure 34c,d,g,h) was used to identify correctly the puncta along the nerve cord. The sorted animals were manually analyzed by behavior to verify the sorting accuracy. Figure 34i shows sorting results from the two outlets; the overall sorting accuracy was 97.7% for a single round. The false positive rate and the false negative rate were 0.2% and 6.6% respectively.

The sorting of strains CZ5261 and CZ5264 relies on determining the locations of GFP along the nerve cord. This is done using the same methods as mentioned earlier, namely compressing the z-stack to the xy plane and then convolving it with a matrix to accentuate the puncta. A simple threshold is subsequently applied to the image to locate the puncta. This work focused on the sorting of wild-type animals from previously determined mutants based on a decision boundary created by classifying both populations using a variety of features (morphological and not). To do this, features that might provide information on differences between the mutant and wild-type populations were selected. Some of these selected features are shown in Figure 35. With this information, it was relatively easy to create a decision boundary for classifying and sorting.

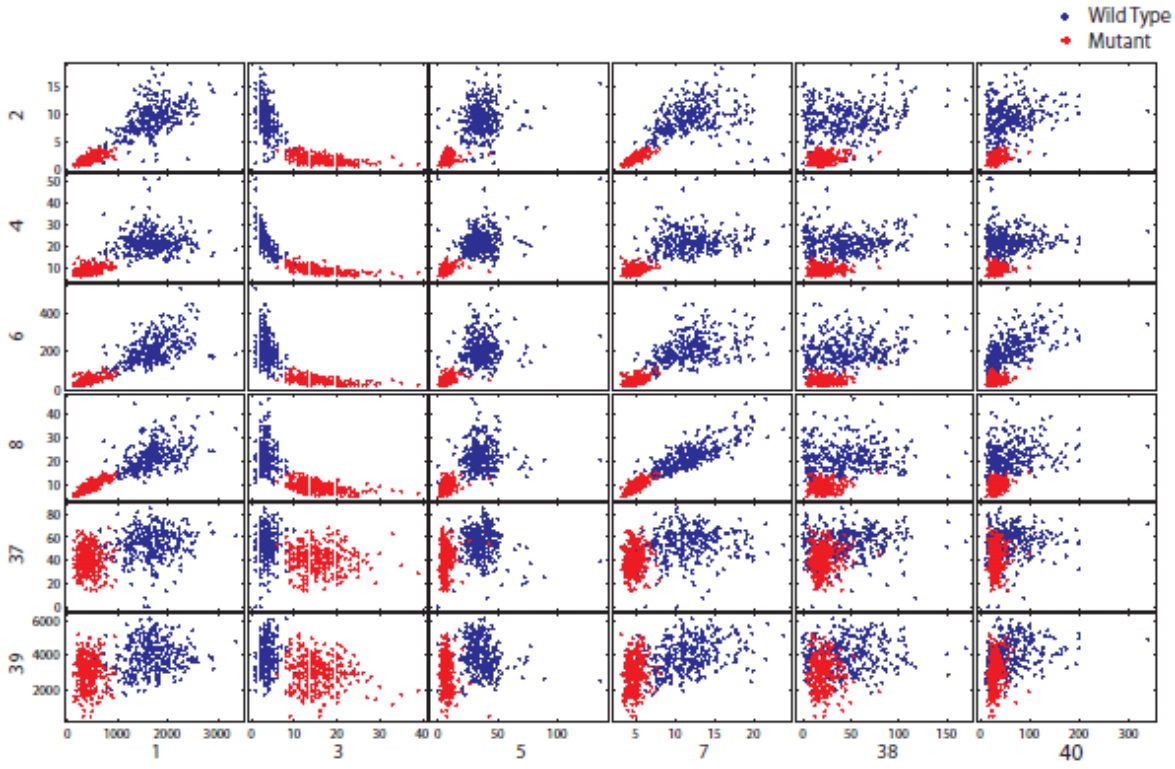



**Figure 35: Graphs showing the distribution of a variety of features used to classify animals as either wild-type or mutant. (a) A variety of selected features that show different degrees of distribution and efficacy in discriminating between classes. Features 2, 4, 6, 8, 37, 39 correspond to std of minor axis length, mean of minor axis length, std major axis length, mean orientation, mean ellipticity. Features 1, 3, 5, 7, 38, 40 correspond to total area, number of puncta, mean area, std of area, std orientation and std ellipticity.**

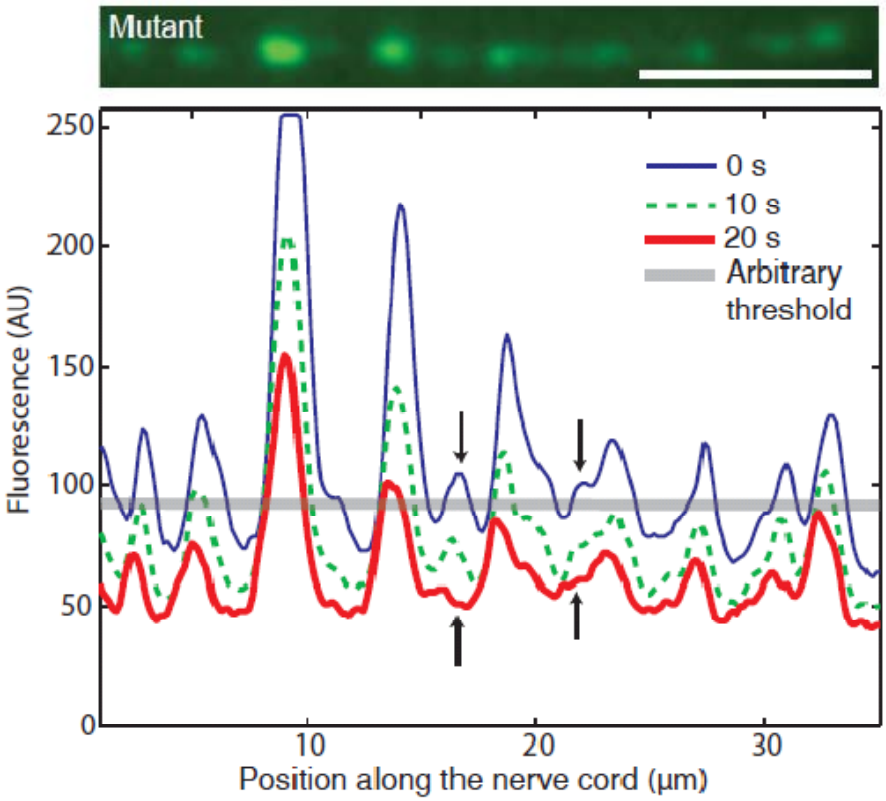



**Figure 36: Top: puncta structures of the nerve cord in a mutant animal before significant photobleaching. Bottom: quantification of puncta fluorescence from line scans as photobleaching occurs. The use of a threshold to determine the number of puncta structures could result in puncta being miscounted depending on the extent of photobleaching. Arrows point to two small punctal structures that would not be identified after deliberate photobleaching for 20 seconds.**

Quantitative analysis of reporter intensity can be performed in the images that are recorded. Figure 36 shows one such example of the *juIs198* marker along the nerve cord of the mutant. The number of puncta and size distribution of the puncta can be automatically analyzed. By implementing a wide variety of classification features (such as average intensity, distance between features, and a wide range of morphological characteristics), it is possible to perform sorting based on any number of features that differ from those of wild-type. The quantitative approach has another advantage: the analysis of reporter intensity often can be strongly influenced by photobleaching of markers and inconsistency of handling between samples(131, 137). Figure 36 shows the effect of deliberate photobleaching on the *juIs198* marker. The system minimizes photobleaching by avoiding manual focusing; data with minimal bleaching

(e.g. 30 msec as in this experiment) can be recorded and used for subsequent analysis. In addition, all animals receive the same handling in the automated microsystem and images are analyzed uniformly using the same criteria, reducing noises and biases introduced by manual operations.

## PERFORMANCE OF SYSTEMS LEVEL COMPONENTS

The external systems components discussed in this chapter were central to the operation and development of the microfluidic system. Engineering computer control for the on-chip valves allowed automation of the device operation, and the creation of a closed loop control system that is discussed in following chapters. The time spent packaging the controls and regulators into a single unit allowed other labs with limited or no experience with microfluidics to use the microfluidic devices discussed in this work. Currently two biology labs are using these system-level components to perform imaging and conduct screens.

**Table 4: A summary of the systems designed that allowed computerized control and automated operation of the microfluidic device for an extended period of time. These tools allowed the microfluidic device to perform simple automated sorting.**

| **Problem** | **Solution** |
|---|---|
| Computerized control of valves | • Amplify the output from a computer controlled DAQ board to power solenoid valves |
| Injection of animal suspension | • A pressure injection container that would allow reliable injection of a mixed solution of animals and buffer |
| Immobilization through cooling | • Created a closed loop cooling system<br>• Programmable temperature control down to 15C<br>• Low, constant flow rate using a peristaltic pump to ensure constant heat removal from device |

# CHAPTER 4: AN AUTONOMOUS SYSTEM FOR EXTENDED IMAGING AND SCREENING

The system discussed in a Chapter 3 for automated sorting was primarily used as a vehicle to prove that automated sorting was technically feasible. This system, however, lacked error handling and thus required nearly constant supervision during operation. A simple failure would cause sorting to cease, and thus while the automation was technically interesting, it rendered the "automated" sorting somewhat irrelevant to actual biological work. The external macroscale control components needed to be integrated with the microfluidic device to allow closed-loop control. The closed loop control required error handling routines to reduce the need for operator intervention, and optimizing the operational sequence was required to minimize the amount of time spent per animal. A comprehensive framework for handling errors robustly so that extended screening could occur without babysitting was developed. The general framework for system operation is shown in Figure 37.

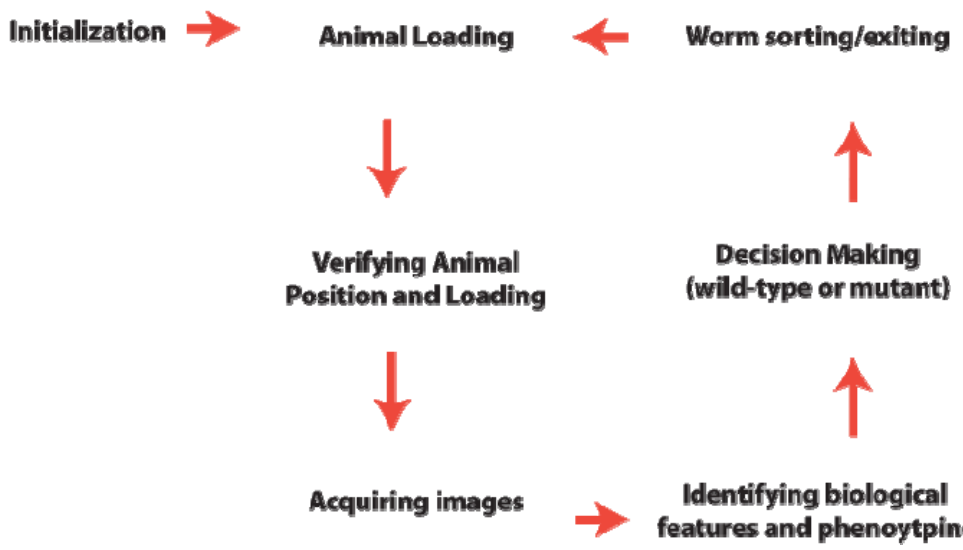



**Figure 37: System schematic as a whole. Following initiation of the system, it enters a closed loop wherein animals are loaded, imaged and screened without interruption. The computer vision portions (grouped in the lower right) will be addressed in the following chapters.**

The critical components of this framework will be dissected and addressed in this chapter. In particular, this will focus on optimizing the system operation and flow in order to minimize bottlenecks and unnecessary operations as well as the creation of error handling routines.

## SYSTEMIC OPTIMIZATION CONSIDERATIONS

One of the critical elements of the system design is the throughput. Doing genome-wide forward genetic screens quickly becomes a relatively daunting numbers game, and one where throughput quickly becomes a significant. During the preliminary demonstrations on using an automated system to sort different strains, the throughput was significantly affected by type of sorting and resolution required (Chapter 4). For a high magnification sorting looking for features that were only present in either the head or tail, throughput was limited to ~60 animals per hour. Performing genome scale screens with a throughput of this level, would not be feasible. To increase the throughput several fold, I worked on identifying the bottlenecks and eliminating them.

### IDENTIFYING BOTTLENECKS

The third-generation doubled the throughput by improving the consistency of device operation and insuring that animals were released consistently. Because of the decreased failure rate and the reduced likelihood of clogging or animals exiting in a poor manner, the throughput was increased two-fold. Because the animals could be positioned head or tail first and could also be different sizes, the original procedure was to acquire dense z-stacks along the entire length of the animal. This ensured that the features of interest were imaged regardless of how the animal was positioned or oriented. A breakdown of the time spent at each of the steps is given below in Table 5. Although the small thermal mass of the animals allows them to be cooled down extremely quickly, occasionally animals move slightly or fluidic capacitance within the device causes small fluid flows within the first couple seconds. Because of this, a 2

second delay is used prior to imaging the animals. Additionally, it is assumed that the images from roughly 10% of the animals will be unusable. Poor image quality results when the synapses are further from the objective and thus distorted by the animal body, or when synapses are touching the PDMS sidewalls and the difference in the index of refraction causes reflections and distortions. Using these assumptions, the processing times in Table 1 were derived.

**Table 5: The original imaging procedure. Dense z-stacks were acquired along the entire length of the worm.**

| Step | Time |
|---|---|
| Wait for worm to load | 1s |
| Pause to ensure immobilization | 2s |
| Acquiring 60 um stack with 2 um steps | 4s |
| Moving to next position on the worm | 1s |
| Acquiring 60 um stack with 2 um steps | 4s |
| Repeat moving, imaging 2 additional times | 10s |
| Image processing to identify synapses | 2s |
| Phenotyping and decision making | 0.5s |
| Sorting | 0.5s |
| Subtotal | 25s |
| Total (assuming 90% good images) | 27.8s |

Although this allows for significantly faster sorting based on device optimizations, it was still limited to ~120 worms per hour. The bulk of the processing time (15s) is devoted to acquiring images at multiple locations along the length of the animal. Because, statistically, any animal is just as likely, or unlikely to contain an interesting mutation, it is unnecessary to image each animal that goes through the device. Therefore, it is more important to increase the throughput, and this can be disregarding animals that are in the wrong orientation. Under this assumption, 50% of the animals enter in the wrong orientation, and only 80% of the worms in the correct orientation have sufficiently good image quality. These are both realistic, but conservative estimations. The results of these assumptions can be seen in

Table 6.

**Table 6: Optimized screening protocol. Rather than focusing on imaging all the animals, only those in the correct orientation are phenotyped.**

| **Step** | **Time** |
|---|---|
| Wait for worm to load | 1s |
| Pause to ensure immobilization | 2s |
| Acquiring 60 um stack with 2 um steps | 4s |
| Image processing to identify synapses | 2s |
| Phenotyping and decision making | 0.5s |
| Sorting | 0.5s |
| Subtotal | 10s |
| Assuming 50% of worms are correct orientation | 20s |
| **Total** (80% correct size and good images) | **25s** |

This resulted in a significant, but not dramatic 10% increase in throughput. Although each animals is processed considerably faster, a significant amount of time is wasted with each of the animals that is in the wrong orientation. Acquiring a dense z-stack and attempting to locate the synapses is a considerable drain on the throughput. The ideal situation is presented in

Table 7. In this scenario, following worm loading the animals would be classified as the correct orientation, or the incorrect orientation. For this to work properly, a feature set and classifier would have to be created that would be capable of discriminating between the two classes (heat and tail) given only the image plane and not a complete z-stack. This proved challenging the image plane could vary +/- 20 microns during the course of the screening. The development of this classifier is discussed at length in Chapter 5. These optimizations in the screening protocol and the device allowed for a three-fold increase (>200 animals per hour) in throughput relative to the first generation system.

**Table 7: The optimized screening protocol. Animals with the wrong orientation are identified as such and discarded before acquiring a dense z-stack. This increases the throughput 50% compared to the original protocol.**

| **Step** | **Time** |
|---|---|
| Wait for worm to load | 1s |
| Pause to ensure immobilization | 1s |
| Check Head/Tail orientation | .5s |
| Sorting | .5s |
| Wait for worm to load | 1s |
| Pause to ensure immobilization | 1s |
| Check Head/Tail orientation | .5s |
| Pause to ensure immobilization | 1s |
| Acquiring 60 um stack with 2 um steps | 4s |
| Image processing to identify synapses | 2s |
| Phenotyping and decision making | .5s |
| Sorting | 1s |
| Subtotal | 14 |
| **Total** (90% correct size and good images) | **15.5s** |

## HANDLING ERRORS ASSOCIATED WITH AUTOMATED SCREENING

One of the critical components of creating a robust system that can screen thousands of animals for hours at a time was the development of error handling routines. Through careful design and optimization of the device and screening protocols, it is possible to reduce the number of errors during normal operation, but it is not possible to eliminate them. A system for extended screening, therefore, relies on extensive error handling to deal with cases for when system performance is non-optimal.

1) Flow errors. Grouped into these is when worms take too long to arrive or exit.
2) Orientation and image quality errors.
3) Computational errors.

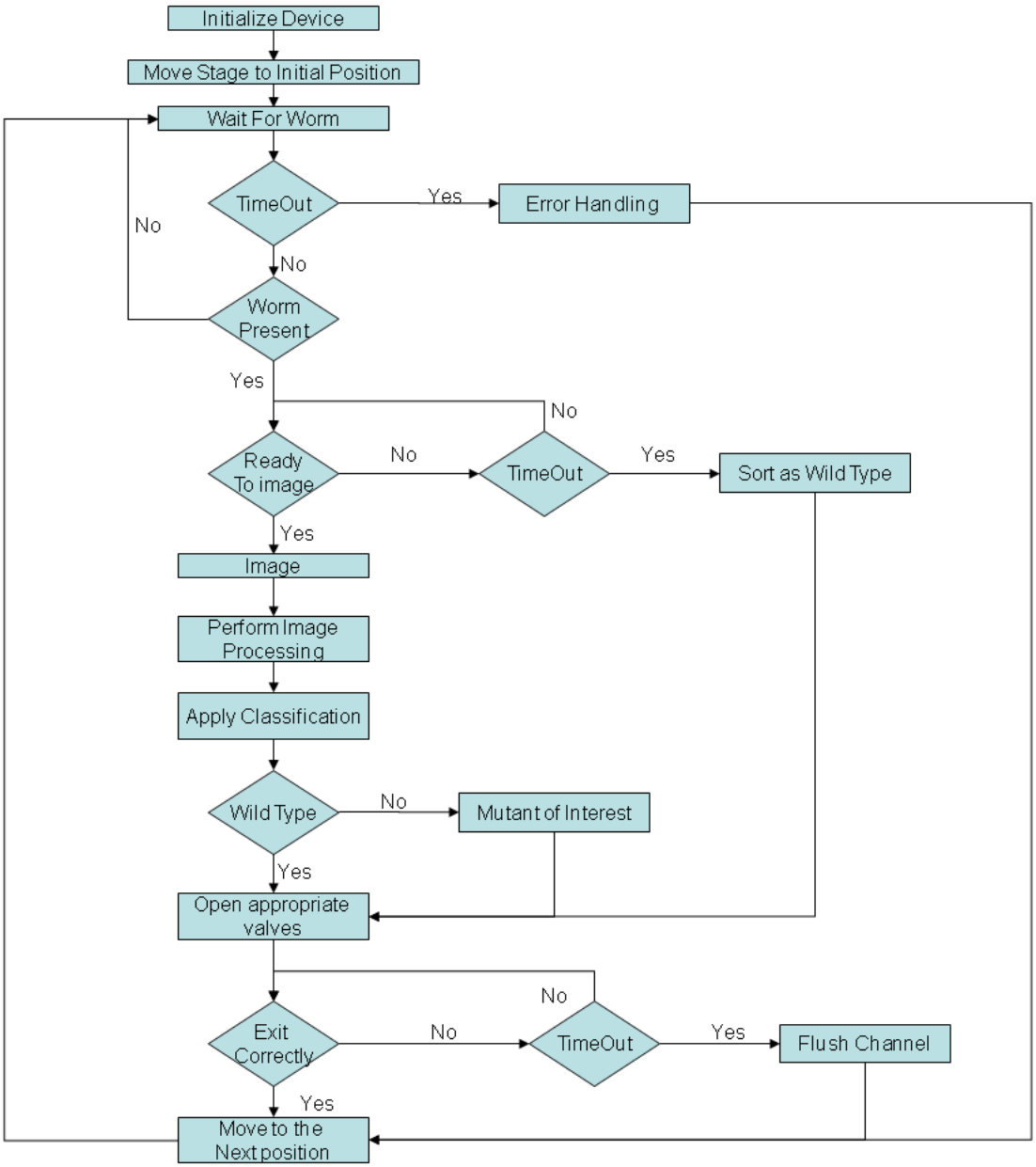



**Figure 38: Detailed flow diagram of the system operation and the appropriate error handling.**

## LOADING AND EXITING ERRORS

There are three primary sources of loading and exiting errors.

1) Clogging of the device due to debris present in the fluid
2) Loading errors because animals have folded over and are blocking the entrance
3) Exiting error because of the size of the animal or clogging.
4) Low density of animals or lack of animals in injection fluid

Because of the limited field of view during standard device operation, although there are four different scenarios (above) that could cause fluid flow problems, there are only inputs that can be detected: the presence of a worm, or the absence of a worm. These limited observations imply that the error handling begins with the most benign assumption, and attempts to correct assuming that error. If this fails to correct the problem, it then proceeds to the next error correction attempt. Finally, if the error cannot be corrected, the user is notified by text message that intervention is required. This correction framework is shown in the Figure 39, below.

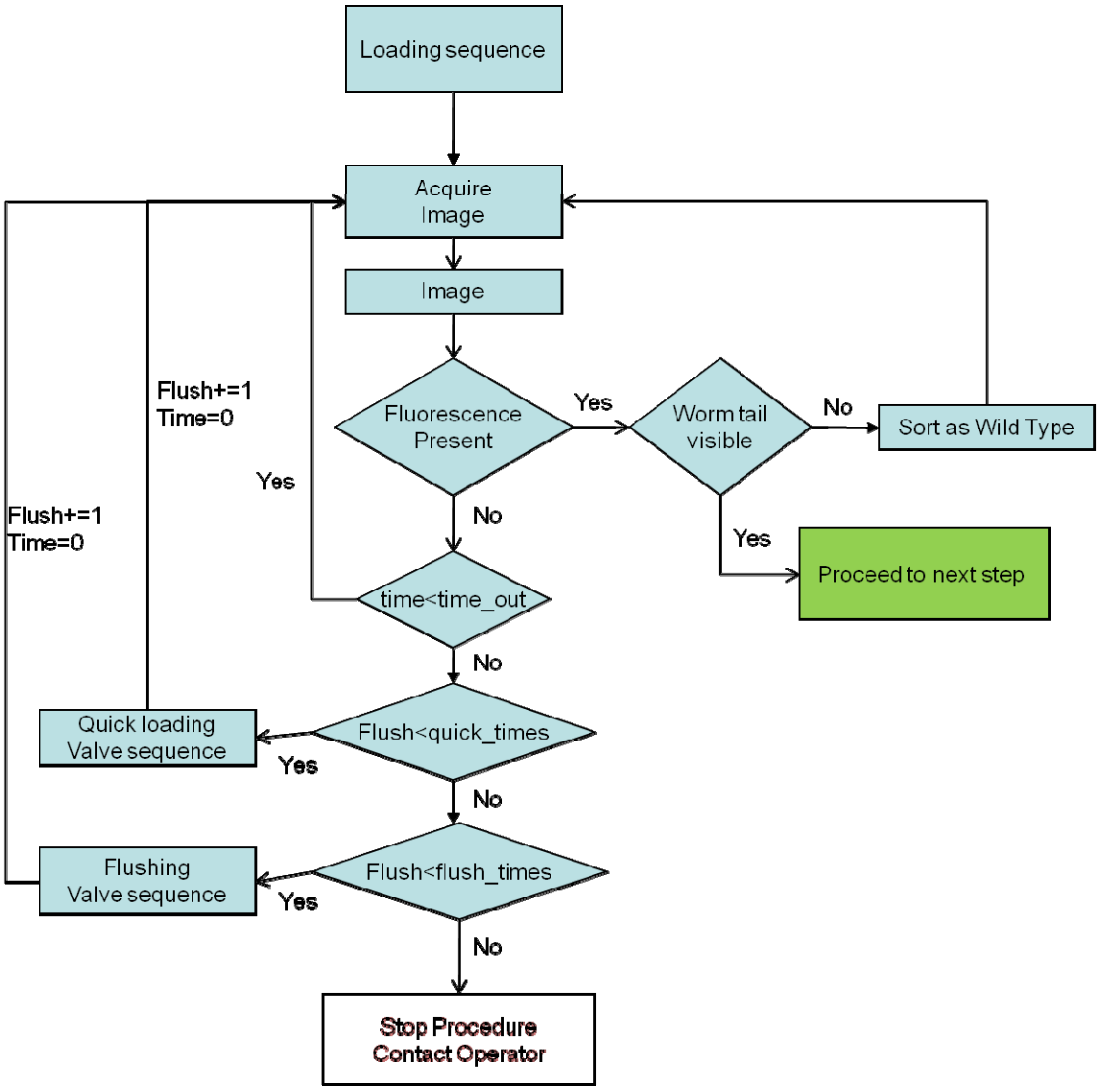



Figure 39: Flow diagram for animal loading including the error handling

For animal loading, the most trivial is that the animal loading distribution has shifted slightly, and that animals are further up the device than assumed. For the first two timeouts, it is assumed that animals are taking longer to load, and thus the valves are rapidly (<100 ms) actuated to allow faster fluid flow through the device and allow rapid animal loading. If this fails to solve the problem, and additional timeout occurs, it is assumed that an animal could be blocking the entrance. Occasionally, rather than entering with its longitudinal axis aligned with the imaging channel, a worm will be positioned perpendicular to the

channel. The fluid flow and pressure differential cause the worm to fold, but the cross-sectional area is significantly larger than the channel, and will prevent it from entering. In this case, the flushing channel is opened briefly. This serves two purposes as it pushes anything trapped in the imaging channel out, and because the flushing pressure is greater than the injection pressure, a small amount of fluid flows into the main chamber, dislodging the folded animal. The most serious case is the clogging of the device by debris present in the injection fluid. The injection fluid is typically filtered through a 0.22 micron filter prior to use, but the solution containing animals cannot be filtered. The standard procedure is to rinse the animals repeatedly, wait for them to sediment and then exchange the solution repeatedly. Even a single rigid particle or piece of dust has the potential to clog the device. In the case of a suspected clog, the flushing fluid is turned on for an extended period of time in an attempt to dislodge the object. If this fails, it could be either due to a clog, or there might be no additional animals to be screened. In this case, the operator is contacted for physical debugging and troubleshooting.

For animal exiting, the process is extremely similar to animal loading. The primary exception is that very rarely is clogging a concern. Error handling is included for this in the situation that it occurs, but only twice has it occurred out of the many devices and experiments that have been formed. Typically, if the system reaches the exiting portion, it is straightforward to remove the dust or animal. If an animal fails to exit normally, and a timeout occurs, a flushing/sorting sequence is initiated and repeated several times until the worm exits. If it fails to exit within several rounds of this, the user is notified of the potential clog and operations are stopped.

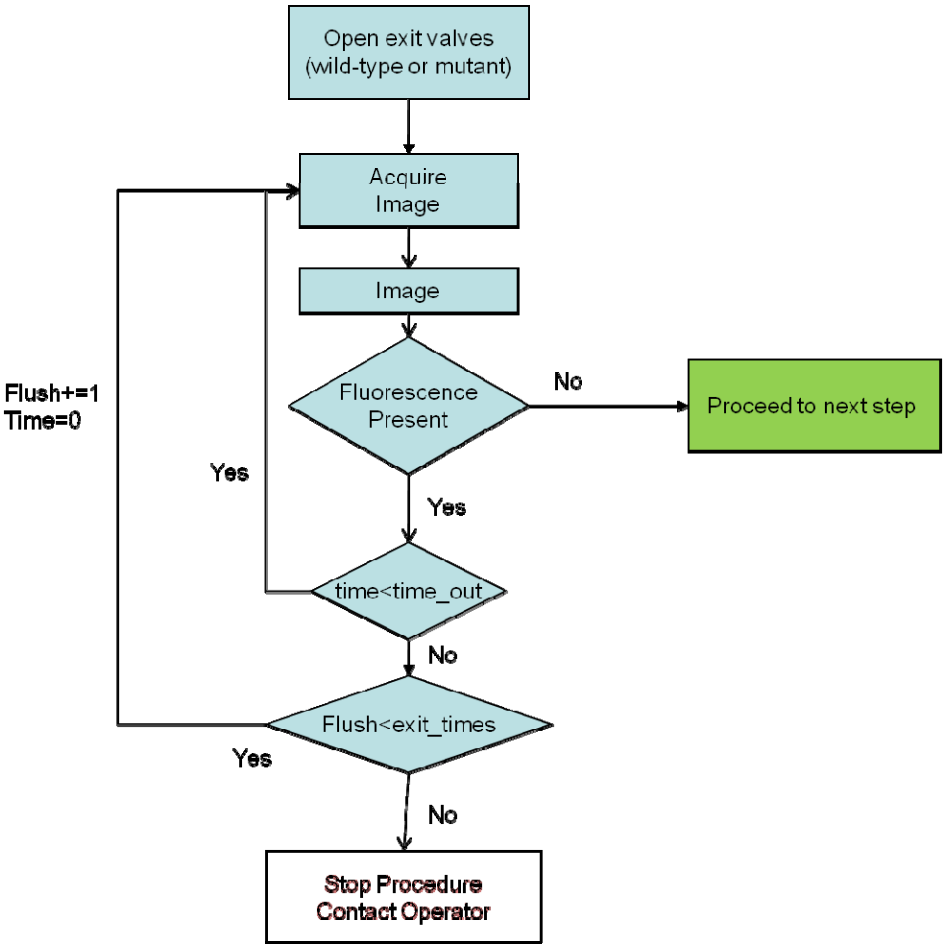



**Figure 40: Flow diagram showing the procedure for animal exiting from the device**

## HEAD/TAIL/NON-TAIL IDENTIFICATION

The head and tail identification was one of the critical elements that allowed screening throughput to approach levels that would make genome-wide screens a reasonable proposition. As was discussed earlier in this chapter, correctly identifying whether the tail was in the field of view, prior to acquiring a z-stack and additional computations, would result in a nearly 60% increase in throughput relative to the baseline. This is a potentially tricky problem for several reasons:

1) Intra-class variance is high. Because the populations being screened have been mutagenized, there is significant variation in levels of autofluorescence. Additionally, the size of the animals varies significantly.
2) The processing must be extremely fast. In order to provide a significant increase in throughput, the feature extraction and classification must occur rapidly (in <1/2s). Because feature extraction and classification must be performed for every animal, any increase in processing time creates a significant burden on the system.
3) Only the focal plane that is in focus at the time of animal entry can be used for feature extraction. Because animals can enter in many different rotational orientations, and the focal plane of the objective relative to the height of the imaging channel will vary between or even during experiments, the image will differ significantly. The focal portion of the animal in focus can vary by +/- 30 microns.

The system implemented to classify the image holistically used a combination of features and for specific features of interest, and histogram derived features. In particular, some of the most helpful features were in terms of the ratio of the Green:Red signals present in the image. Fortunately, the strains used in the imaging had an mCherry reporter in the head of the animal, and a GFP signal solely present in the tail. Examples of the significant degree of variation found in the images are shown in Figure 41.

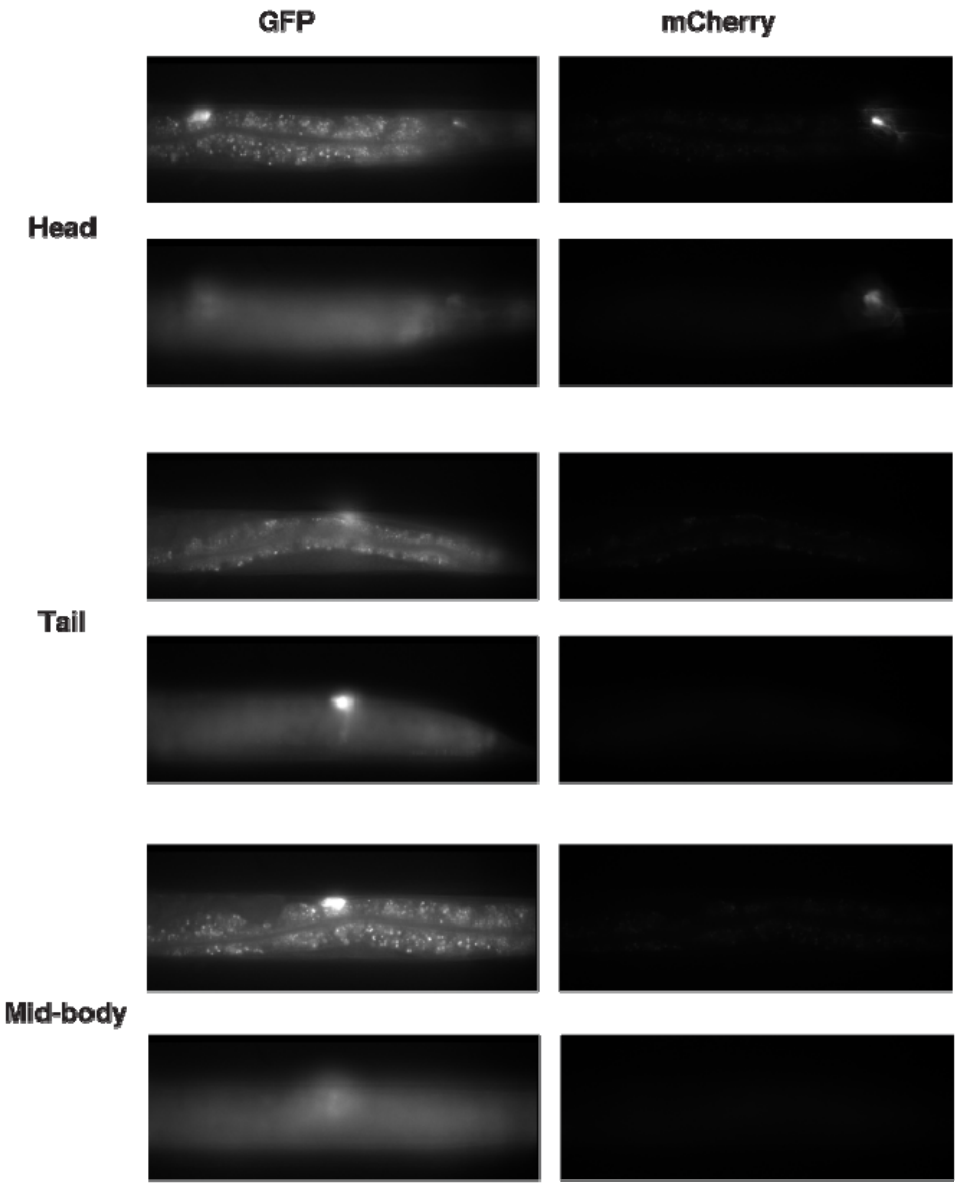



**Figure 41: Examples of the degree of image variation in the three different classes of animals. Head (top) appear extremely similar in the GFP channel, but have a marker present in the mCherry channel. Tail (middle) can occupy anywhere from a third to the majority of the image channel, and have little fluorescence in the mCherry channel. Mid-body or partial tail images (bottom) appear extremely similar to the tail images.**

The features extracted from the images were based on heuristic principles. The image plane provides two sets of information: a red-channel image and a green-channel image. Knowing the offset between the two images (determined prior to screening), an additional image is extracted: a Green:Red channel ratio

image. Because the raw intensity image in either channel can be affected by camera settings, animal age or any number of other uncontrollable factors, it is helpful to have an image that corrects for changes in imaging conditions that would cause in increase in fluorescence across all spectral channels. For each of these three extracted images, the same set of features is extracted. These features include specific features regarding the distribution of the intensities within the image such as the mean, standard deviation and information about the skewness. A histogram of both the image and the gradient of the image are then extracted. These histograms are extracted over different ranges for each of the three feature images. Histogram ranges are kept constant between animals rather than using the max/min of each image, as this ensures operational consistency.

These feature sets proved extremely powerful at discriminating between the head and tail cases, but less so at discriminating between classes “2” and “3” (tail or partial tail). This was largely because there was considerable overlap between the two classes in the feature space. An animal that was relatively out of focus and only a partial tail might appear similar to a correctly positioned animal but in-focus. To provide additional features, the image was separated into segments as seen in Figure 42. Additional features were calculated using just these sub-portions of the image to improve the classifier performance. In total, 164 features were calculated per image.


**Figure 42: To compensate for the similarities between the two classes (tail and near-tail), the same set of features was extracted from the sub-portions of the image shown underneath the blue boxes. This significantly improved the classification and increased that actual screening throughput.**

A mixed set of both mutant and wild-type animals were used for the training sets. The images were from z-stacks acquired in a microfluidic device. Because the images during screening could come from a large

range of focal planes, the images from the focal planes +/- 20 microns around the center of the worm were extracted and used as the training sample. The images where the head was clearly within the field of view were identified and labeled class “1”. Images where the tail appeared clearly were identified, and called class “2”. Images containing either the middle of the worm (because it was too long) or a portion of the tail were classified as class “3”. Using multiple classes enabled emphasizing the various classes and providing different training weights. Because the problem of classifying a tail image as a head image, and vice versa, is more problematic than classifying either of them as a mid-body or partial-tail, using three separate classes allows the weights to be tuned to emphasize this. The completed library contains features from 4,466 labeled images.

The classifier was an RBF-kernel SVM(87), and the feature sets were normalized on 0 to 1 scale prior to training. The optimal C and gamma were identified using a brute force grid searching and five-fold cross-validation. They were identified as: $C = 16\ and\ \gamma = 2$. The feature set and the training allowed >99% classification accuracy during training. When applied to a separate dataset from screening that had been labeled in isolation, the system performed extremely well. Head or tail confusion was non-existent, and relatively low tail or partial tail confusion was reported.

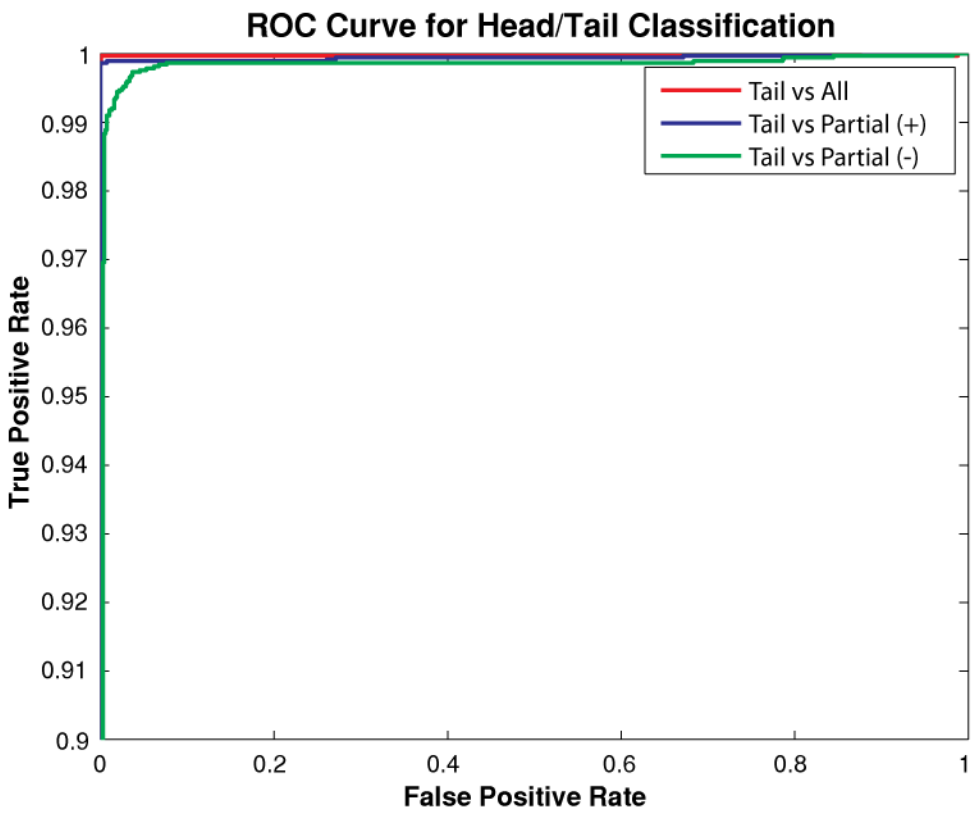



**Figure 43: ROC curves for the head and tail classifier. The y-axis begins at 0.9 to better show the separation between curves. The error rates were determined using 5-fold cross-validation and a grid-search for parameter optimization for each classifier. The use of additional features significantly improved the tail versus partial tail classification.**

In order to measure the performance of the whole image classifier (head/tail/partial tail), performance statistics were calculated. The library of ground-truth labeled images was used for training the classifier for each of the scenarios. To compare the performance increase when the additional features for tail or partial tail were incorporated, an ROC curve was calculated for Tail vs Partial Tail detection with the features (+) and without the features (-). As can be seen in Figure 43, the additional features improve the performance. During each of the cross-validation performance tests a grid-search was performed to determine optimal training parameters for the training. Individual measurements of classifier performance derived from these curves can be seen in Table 8.

**Table 8: Statistics comparing the performance of the head and tail classifier. The classifier performed extremely well with very low numbers of false positives. These statistics were derived from the ROC curve above. AUC is the area under the curve, EER is the equal error rate (when 1-TP = FP), and the True Positive rate when False Positives equal 1% and 0.1%.**

| | AUC | EER | TP @ FP = 1.0% | TP @ FP = 0.1% |
|---|---|---|---|---|
| Tail vs All | 99.97 | 0.07 | 99.97 | 99.92 |
| Tail vs Partial Tail (+) | 99.94 | 0.15 | 99.89 | 99.87 |
| Tail vs Partial Tail (-) | 99.8 | 0.88 | 99.1 | 64.90 |

## BOW FOR HEAD AND TAIL CLASSIFICATION

The idea underlying the bag of words (BoW) method is that invidual descriptors mean more when put in context to the descriptors around it. Prior to the classifier described above, the BOW method(105, 106, 108) was applied to the whole image classification. The idea underlying this was that the head and tail images must be composed of different numbers of word, and that the composition of these images based on these image descriptors could be a strong way of separating the classes. Additionally, because the "words" identified during the classification is done automatically, without any human intervention or labeling, it could be more generalizable. This could allow the same code framework to be applied to completely different fluorescent reporters without the development of new features, merely the acquisition of new labeled images. The image was divided into small patches based on the HOG method(101), and the HOG features were extracted. Clustering to identify potential words was done using k-means. Following the generation of the image "words", the labeled head/tail images were trained using five-fold cross validation. During testing, this method failed to perform near the same level as that described in detail above. This is not ultimately an indictment of the BOW methodology, as numerous changes in imaging conditions and the image library were after this was tested. The library of labeled images tested using the BOW method was significantly smaller than the database ultimately used. This caused the clustering to identify "words" that were descriptive of imaging changes between animals. It is important to have a large database of images for training purposes so that small changes in scale, lighting, or other situational differences don't result in poor clustering, and this could have been the problem during BOW training. Additionally, the images used for BOW testing were acquired using only single color

images. The use of the two-color images acquired with the DualView™, could significantly increase the performance. Detailed analysis and comparison the classification ability of various methods, is however, beyond the scope of this work but would be a fruitful area for future research using the above method as a benchmark.

## PERFORMANCE OF THE AUTOMATED IMAGING AND SCREENING

The development of a closed-loop framework to handle errors and ensure consistent system operation during screening was critical to system performance. By including error handling in a closed loop system, the microfluidic devices could operate for significantly longer before requiring human intervention. The error handling system implemented notifies the operator when the system encounters an error requiring intervention, removing the need for constant supervision during operation. By optimizing the operational sequence the amount of time spent per animal was reduced by almost 50%. The results and the methods for this comprehensive framework is shown in Table 9.

**Table 9: Summary of the methods for closed loop operation of imaging and sorting and the resulting performance.**

| Objective | Method | Result |
|---|---|---|
| Closed loop error handling | • Create individual routines to handle the various types of errors (flow, image, and computational errors)<br>• Notify user in case of errors that require human intervention to eliminate need for babysitting | • Increased the mean time between operator intervention from 15 mins to 4.5 hours (1800% increase) |
| Maximize animal throughput | • Optimize the system operation to reduce time spent per animal and minimize unnecessary operations.<br>• Employ classifier to determine whether animal is head or tail. | • Time per animal reduced from 27.8s to 15.5s.<br>• Throughput increased 80% by process optimization alone |
| Classifier to quickly detect worm position | • Create a classifier based around features from a single image plane.<br>• Use RBF-kernel SVM to classify head, tail, partial-tail | • >99% classification accuracy without requiring z-stack acquisition |

# CHAPTER 5: COMPUTER VISION FOR SYNAPSE IDENTIFICATION AND ANIMAL PHENOTYPING

The development of a computer vision framework to automatically handle images of animals in order to make decisions based on the fluorescent patterns present is a complex and extremely challenging problem, that involves multiple stages. The challenges presented by this problem are largely a result of the challenging biological nature of the problem. Phenotyping and sorting of fluorescent patterns can be a simple task if the fluorescent signals are large and extremely bright relative to the surrounding tissues as is often the case with cellular-scale reporters(57, 66, 138). This provides a high signal to noise ratio, which makes it simple to identify the objects of interest, and then to phenotype. In most of these cases, the signal is bright enough that a simple thresholding is sufficient to separate the objects of interest from the background(57, 66). By comparison, the framework presented here allows the identification of small (subcellular) features that are extremely dim and are present in an extremely noisy environment. Because of the small size, limited fluorescence and high background noise, a multi-stage computer vision framework was required to separate the objects of interest (synapses) from background objects of similar size and intensity. These identified features are then used to phenotype the animals. This represents a significant advance over previous methods of phenotyping *C. elegans(66, 111-113)*.

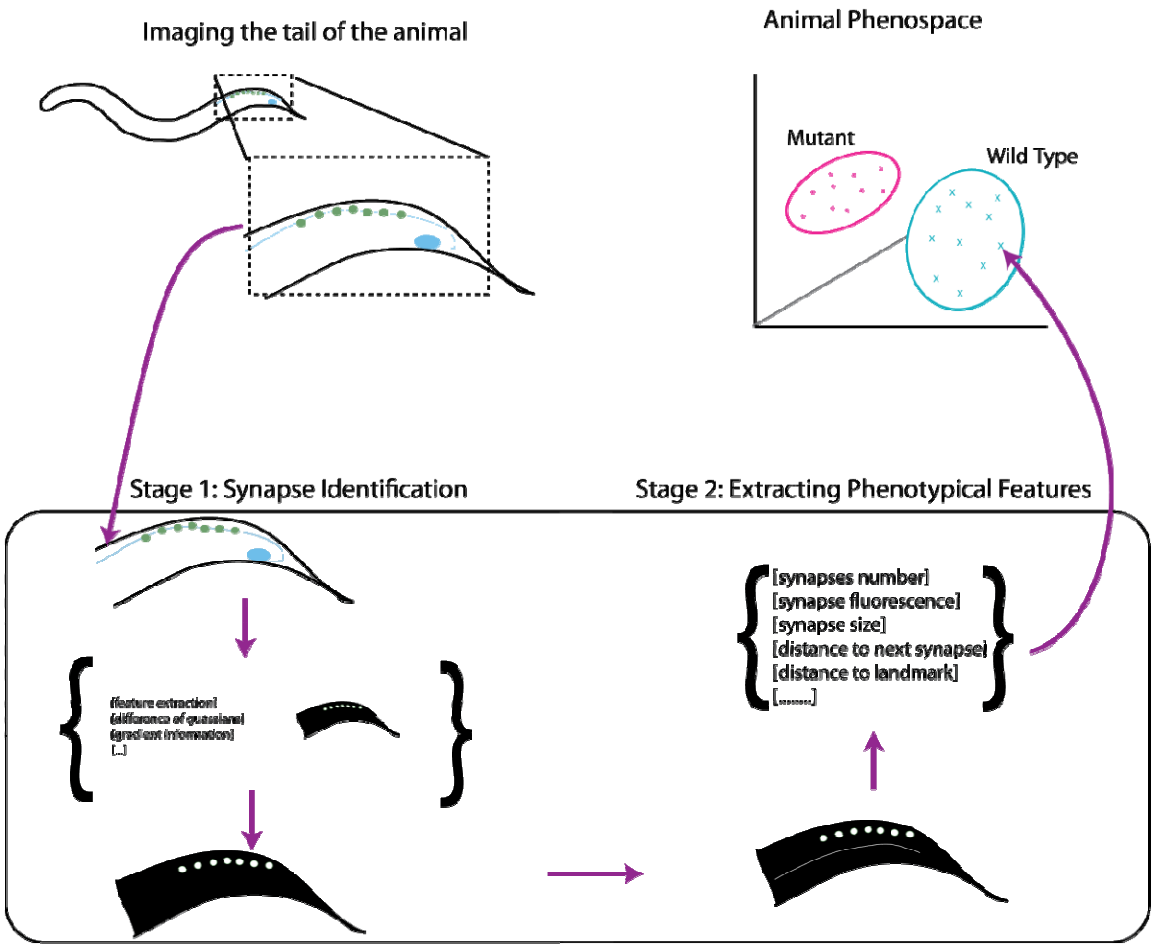



**Figure 44: Schematic showing the two stages of the computer vision process. In the first stage the synapses are identified and in the second stage, the identified synapses and biological landmarks are used to extract specific phenotypical features. These features are then used to classify the animal as a wild-type or a mutant.**

The framework is broken into two sections as seen in Figure 44:

1) A computer vision system to identify the primary objects of interest (synapses). This is a multistage process that separates the synapses from the rest of the image that often includes objects of similar shape and fluorescence intensity.
2) A second stage to identify landmarks in the image and then using the extracted synapses and the landmarks to quantify specific phenotypical features. These features are used to project each

animal (and class of animals) onto the phenotypical space for classification as wild-type or mutants.

## GROUND-TRUTH DICTIONARY

In order to create a trained computer vision system for identifying synapses and screening animals, a ground truth library had to be created. In conventional computer vision problems there are a significant number of curated databases that are used for both training purposes(69, 139, 140), as well as objective measures of performance. Because performance can differ significantly depending on the specific database used for training and testing, it is helpful to have generally available image databases. To create these databases, images were labeled and annotated by hand. This allows other researchers to not only acquire performance measurements that can easily be compared, but also allows them to avoid the painstaking process of acquiring and labeling images by hand. Using curated libraries can eliminate months of work devoted to image acquisition and annotation and allow researchers to focus on the development of optimal computer vision tools(141). For this work, however, image acquisition and ground-truth labeling had to be performed manually.

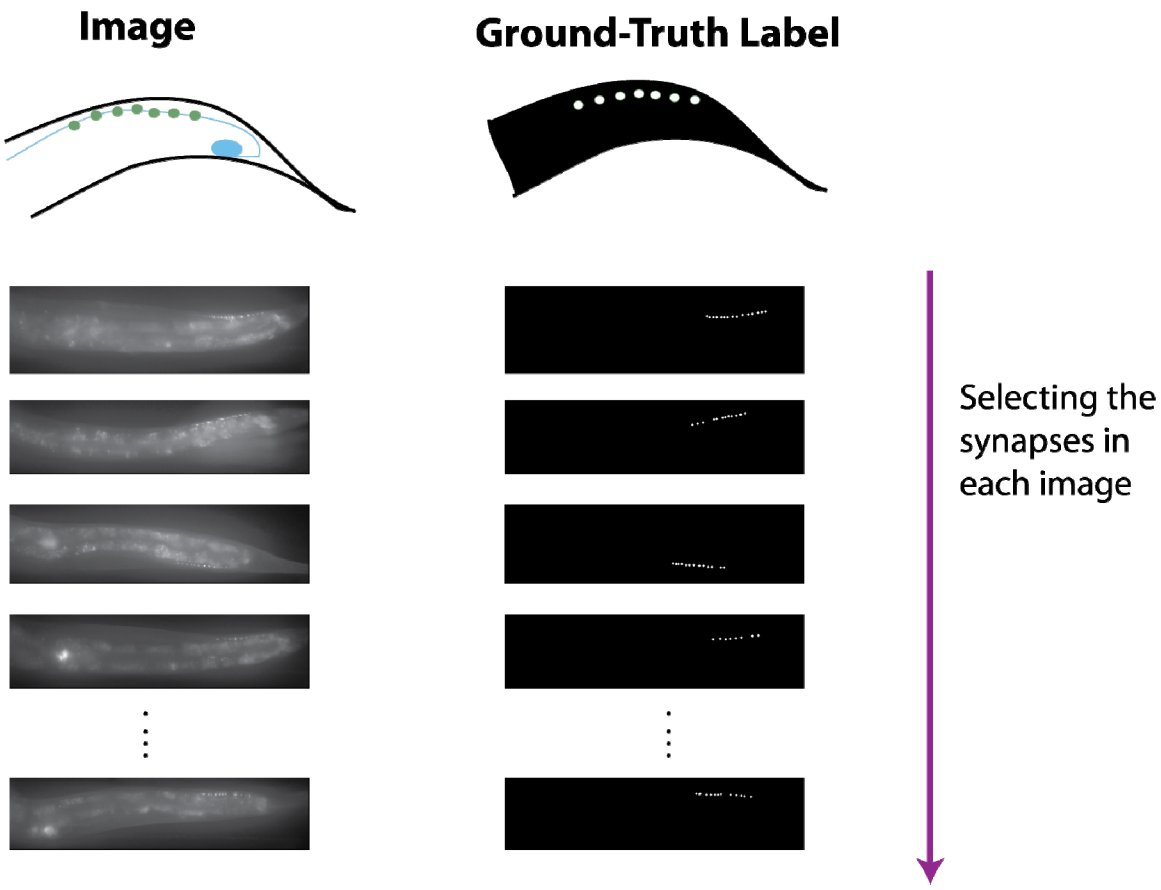



**Figure 45: The creation of a ground-truth library. Z-stacks of numerous strains were acquired in the microfluidic device and saved for future use. This images were later used to create ground-truth libraries by labeling all of the synapses.**

For this project, it was necessary to label the images both holistically, and to identify specific features within each of the images. Each of the images was labeled as head, tail, or mid-body. This labeling was used to help optimize the device operation by rejecting animals that were at the incorrect orientation. In all of the animals that were labeled as tail animals, the synapses were then identified within the image. This labeling process is shown above in Figure 44. The challenge for this, was that if the imaging conditions were ever changed (as happened numerous times), the ground-truth library had to be recreated by imaging and labeling. For example, the original database was created using single color images, and became useless when switching to two-color simultaneous imaging. Additionally, the filters

were changed, and imaging conditions (exposure, etc) were altered several times requiring new ground truth libraries.

## SYNAPSE IDENTIFICATION

The synapse identification is the most challenging of the steps due to the low signal-to-noise ratio (SNR). Examples of the difficulty of the separate the features of interest from the irrelevant features can be seen in Figure 46. Furthermore, the synapse identification is complicated by multiple technical requirements in addition to the low SNR. These problems can be separated into three distinct areas

1) Throughput
    a. Image processing and image recognition must occur in near-real-time. Decision making must occur on-line immediately following the image acquisition, and each additional second reduces the screening throughput. For this work, the objective was to create an image processing system that required less time than the image acquisition.
2) Image properties
    a. As previously mentioned, the images had a low SNR due to the high levels of autofluorescence, small synapses and fat granules with similar sizes and fluorescent intensities. The synapses also have limited numbers of fluorophores and photo-bleach quickly
    b. Animals enter with variations in rotational orientation. Because the animals are surrounded by a cuticle, the body of the animal can act as a lens and imaging through the organism can create distortions in the expression pattern and image.
    c. The small depth of field means that the information of interest (synapses) are often at different focal planes, and thus a pseudo 3-dimensional image must be acquired.
3) Varied animal population

a. Dealing with a mutagenized population of animals introduces a significant degree of variation into the animal appearance. This includes significant variation in size (length and width) as well as changes in levels of autofluorescence or size and distribution of fat granules. Even more significant changes to animal appearance, such as increased numbers of embryos, the appearance of blisters on the cuticle, or even animals that have hatched inside the mother are commonly seen.

These factors imposed constraints on the image processing that had to be considered for each of the steps. Specifically, the variations in the images and image quality make it challenging to correctly separate the correct animals from irrelevant features.

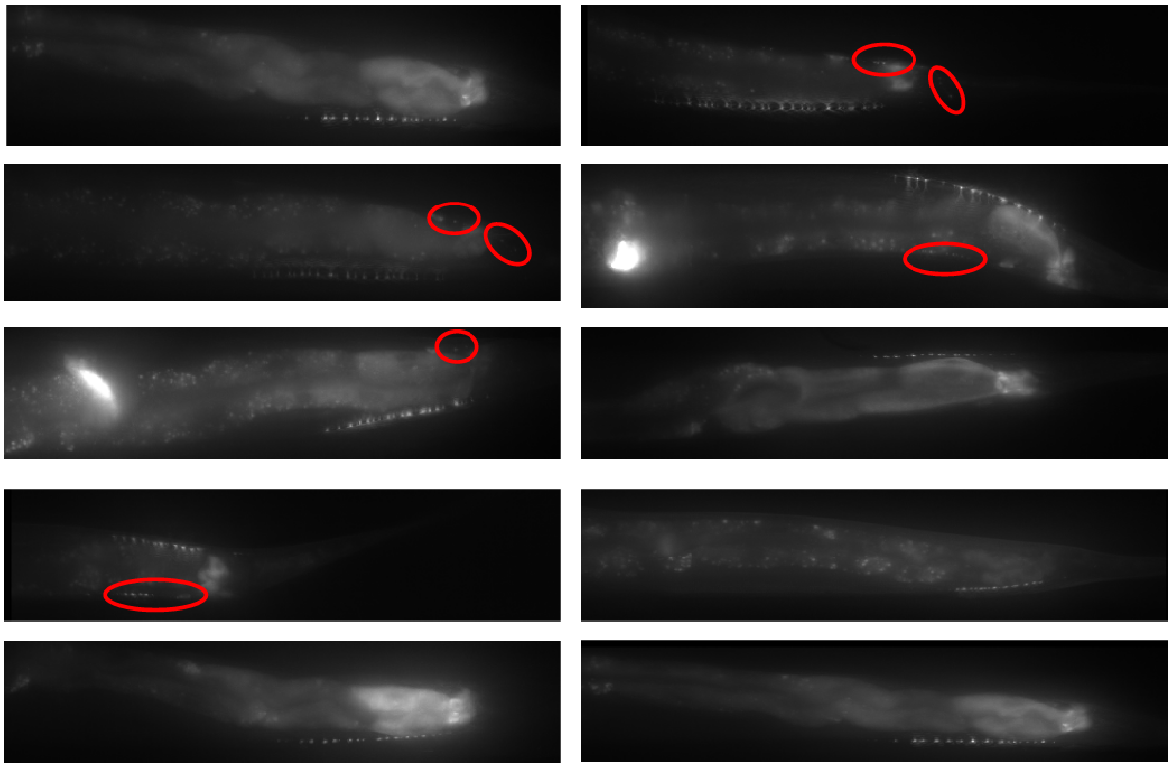


**Figure 46: Images of the focal plane of both wild-type and mutant animals in the device. As can be easily seen by the images, the synapses have a rather low SNR when compared to the surrounding tissues. The synapse that are mislocalized and cause the animal to be identified as mutants of interest are circled in red.**

To deal with these variations, the framework developed performs upfront image processing to reduce the image from a three-dimensional image and to reduce the complexity while maximizing the information present in the image. From these reduced complexity images, local features are extracted based on the nearby pixels. Larger region-based features are then extracted using both the local features and heuristics applied to the image based on knowledge about the animal position and appearance.

The image processing was ultimately separated into three stages:

1) Upfront image processing. This served to reduce the dimensionality of the images while preserving the majority of relevant information. Additionally, the image was separated based on the spectral filter and aligned.
2) Rapid local feature filtering. This was designed to extract features from small neighborhoods surrounding each pixel, and then to use a quick linear-SVM to identify potential synapse locations
3) Regional feature extraction. Using both features extracted from larger regions and the potential synapse locations and relative positions, a second layer of features were created. To prevent this from becoming a bottleneck, only the locations with a high probability of being actual synapses from the previous step were analyzed using an RBF-kernel SVM.

This process and the various stages in the schematic are shown in Figure 47.

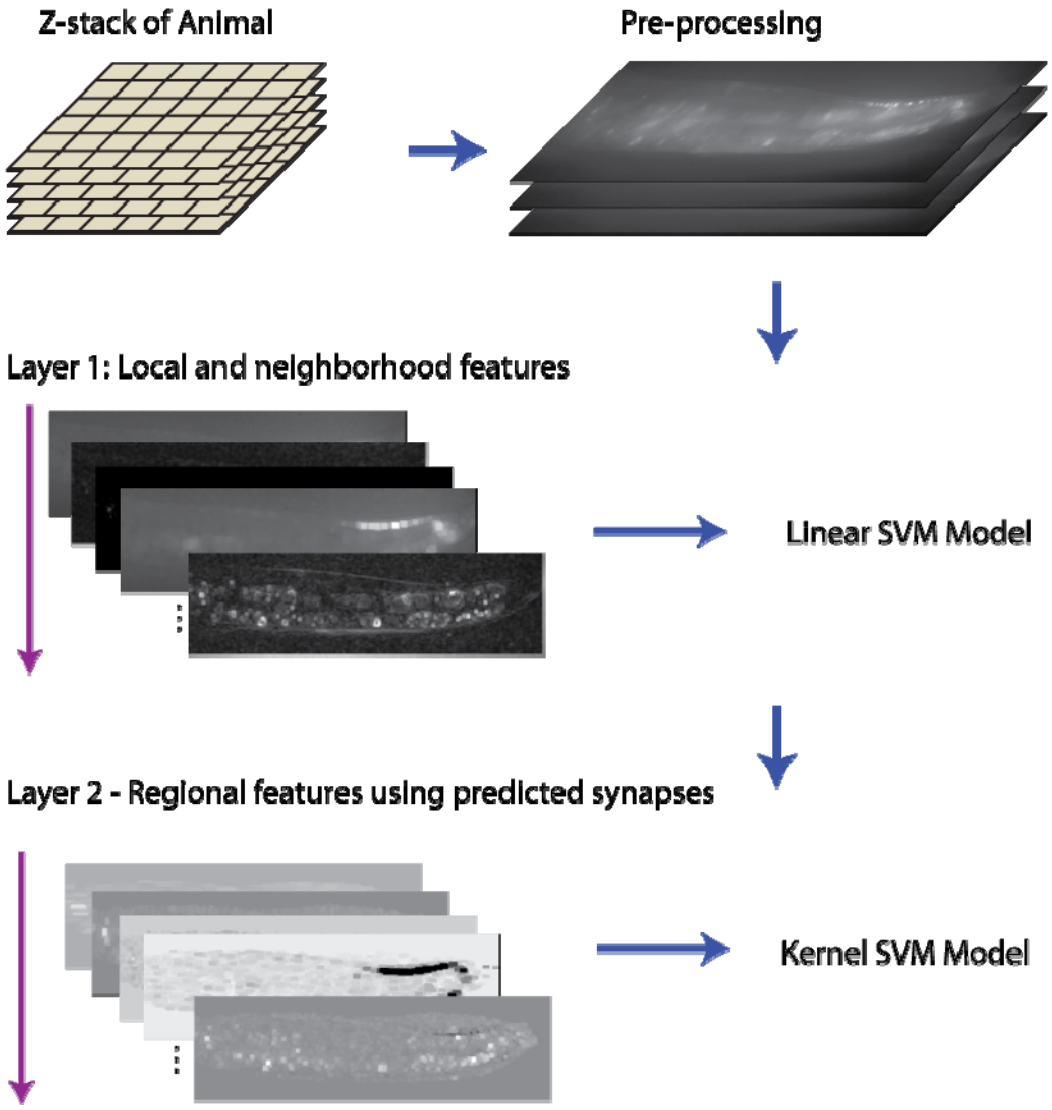



**Figure 47: The three image processing stages for synapse identification. First, the acquired z-stack is processed to reduce the dimensionality and increase the SNR. Then, local features are used to identify probable synapse locations in a rapid manner. Thirdly, the probable synapses and the original post-processing image are used to extract regional features and identify the actual synapse locations.**

## UPFRONT IMAGE PROCESSING

During the image acquisition stage, images are acquired at multiple focal planes at regular steps along the z-direction of the animal. For most of these experiments, images were acquired at 1.5 micron steps over 60 microns. This ensured that regardless of the original position of the objective, and the orientation of

the animal, the objects of interest (synapses) would be in-focus. Although the synapses tend to be relatively in-plane, they are often present in multiple images. This occurs especially frequently when synapses are on opposing sides of the animal. In these cases, even a small rotational change in the animal can result in synapses being in-focus ten microns apart from one-another. This meant that the image processing could not be confined to a single image plane.

The primary challenge was in dealing with a massive amount of information and identifying the synapses within a few seconds. Images were acquired at a 640X480 resolution, and with 40 image planes, this resulted in 12 million pixels. Processing these in a volumetric manner may have allowed for optimal performance, but it was prohibitively computationally intensive. This volumetric image was compressed into a 2-dimensional image by taking the maximum projection at each x-y point along the z-direction. This process can be seen in Figure 48.

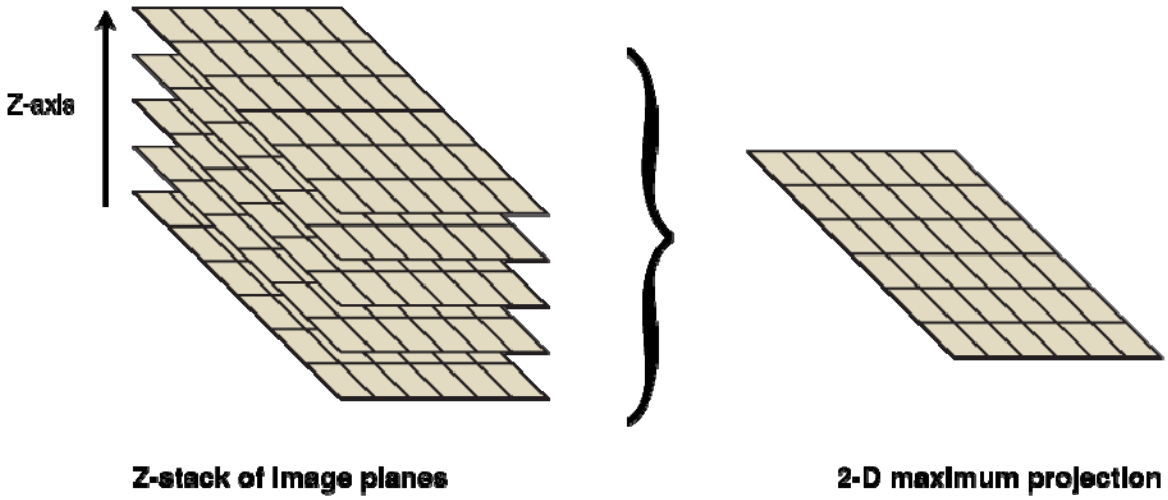



**Figure 48: The conversion of a pseudo 3-D image into a 2-D image. The acquired image is a z-stack of image planes at different z-locations in the animal. The maximum intensity value at each x-y location was used to project the image down to a two-dimensional image space.**

The conversion to a 2-D image using maximum projection reduced the number of pixels required for high level feature extraction, to a much more manageable 300 thousand. Several alternative projection

methods such as mean projection, or standard-deviation projection were tried, but none of these provided improved performance.

Following projection, the flattened image is separated into three distinct images, each of which will be used for feature extraction. The use of the DualView™ makes it possible to obtain simultaneous two-color images of the same focal plane (Chapter 3). Thus, the top half of the flattened image shows the fluorescence in the mCherry channel, while the lower portion shows fluorescence in the GFP channel. Additionally, the imaging process of the DualView™ creates an offset between the different spectral image channels that can vary significantly between individual imaging runs. To correct for this, the two separate images are aligned. Before each screen or imaging run, images are acquired the brightfield mode and separated into a top and bottom image. Using the normalized cross-correlation, the offset between the two images is determined and used for that screening run. Because the absolute intensities can vary, for example increasing across both the red and green channels, a third image is created using the ratio between the green and red channels. This process is seen in Figure 49.

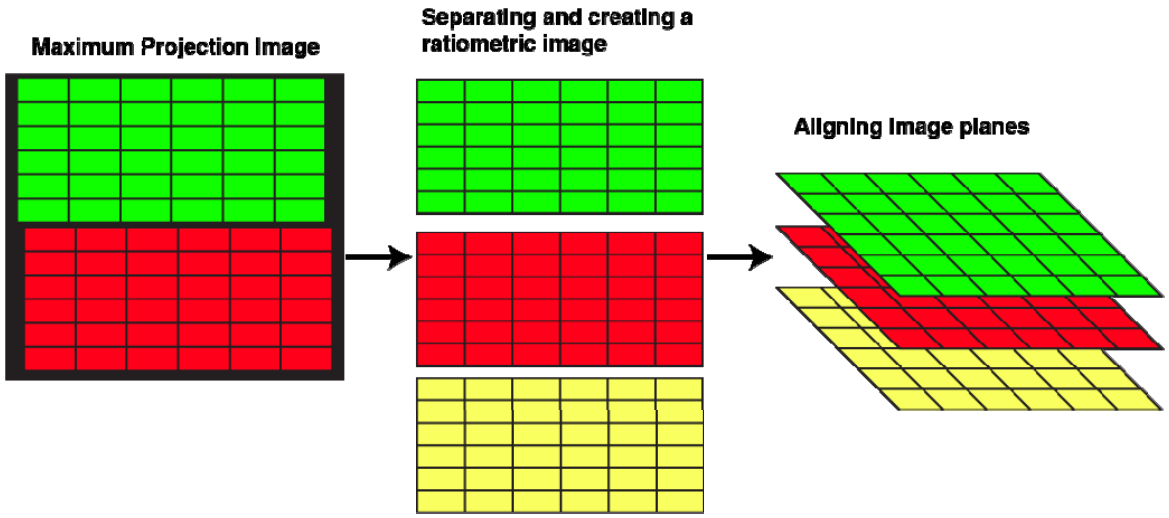



**Figure 49: The transformation of the maximum projection into the separate images. (left) the initial maximum projection image where the green and red images are separated into different regions on the camera. (middle) the identified and separated green and red images are aligned and used to create the ratio-metric image on the bottom. (right) the three images are finally aligned into a multi-layer image.**

This tri-plane image is then used for feature extraction and identification. The use of multiple filters significantly increases the SNR and allows for significantly higher performance, although it increases the feature extraction time accordingly.

## LOCAL FEATURES

The local features are extracted based on small bounding regions. These features were designed to be calculated in an extremely rapid manner and to accentuate the differences between the small synaptic fluorescent expression pattern and the surrounding tissues. Synapses are small regions of bright fluorescence, surrounded by regions of relatively dimmer fluorescence. Furthermore, the ratio of the fluorescence in the green relative to the red images is higher than in the autofluorescence. The majority of these features were calculated in neighborhoods varying from [3 3] up to [8 8].

Because the initial transformation created a matrix containing three separate images (green, red, ratio), the set of local features was calculated on each of the individual images. Using the same set of features increases the flexibility of the system. Additional upfront processing to create additional layers could be easily added, and then the same filters applied.

The filters applied included some adapted from the SIFT features(102, 103). These include the local maxima and minima, gradient, local average, median filtering, difference of Gaussians with different Gaussian filters and the local standard deviation of the neighborhood. Applying these filters over different size scales allows the filters to capture the variability between the fat granules and synaptic expression. Furthermore, calculation of these features is rapid, with all 115 features calculated in <700 msec on a 640X480 image using Pentium Core i7 processor with 4GB of RAM. Following computation of the feature space, the resulting feature matrix was (320X480) or 153,600x115 and applying even a rapid linear SVM to classify all these points would be time consuming (requiring >1s). This was optimized by only classifying the pixels where either the GFP or mCherry intensity was greater than the mean intensity. This essentially

served only classify points within the animal. Because Matlab® performs matrix operations extremely rapidly; this optimization was not applied to the feature extraction, but only the classification.

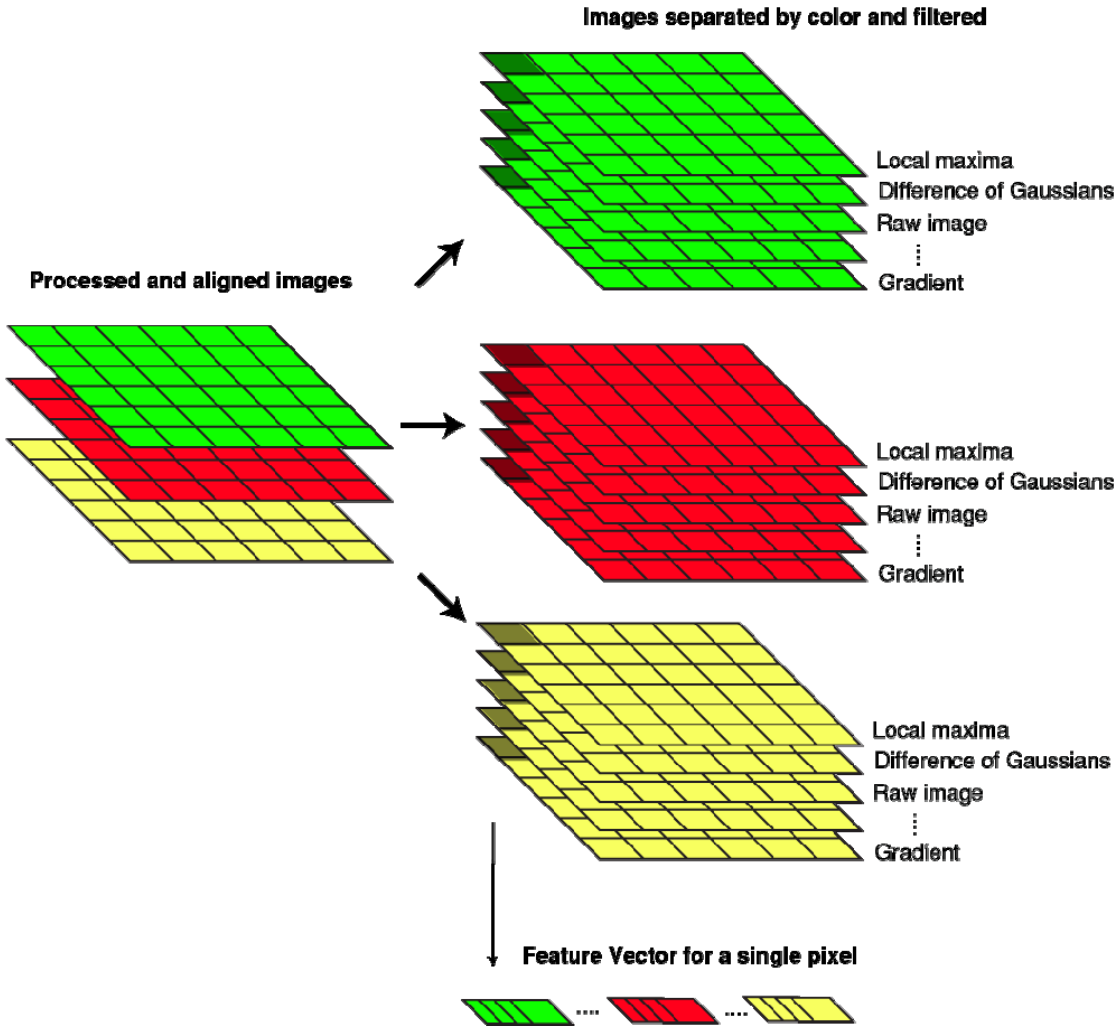



**Figure 50: A schematic showing the process of extracting the local features for the rapid, first pass classification. The initial image containing the green, red and ratio image on the left is separated and used to calculate a sequence of features for each of the images. These features are then used, so that each pixel can be classified using the 115 features extracted for it.**

To train the first layer of the synapse classifier, these features were applied to the ground-truth library discussed in the previous section. The library used for synapse training was composed of 129 images where the synapses present in each were labeled. In each of the labeled image, the full set of features was extracted and then the image was separated into positive and negative training samples. Due to the

size of the synapses relative to the image, the preponderance of pixels (>99.9%) would be negative training samples. To find the points that would be most interesting, and challenging negative training samples, canny edge detection was performed on GFP image and the points in it were broadened. Then, out of these pixels a random selection of 2,000 pixels per image were selected for negative training. This method allowed a variety of training points to be selected from images that were acquired on different days and in different strains to correct for any experimental variability. All of the positively labeled pixels within each image were added to the training dictionary. This process can be seen below in Figure 51.

Using this dictionary, a linear-SVM was trained using liblinear(91). A brute force grid search optimization method using 5-fold cross-validation was employed to determine the optimal training parameters. The synapses were weighted three times more than the non-synapses during training. The optimal cost during the grid search was found to be C=16. Because of the large number of features extracted from the image, a linear SVM was found to perform well. For comparison, an RBF kernel SVM was trained on the same dictionary, and with 5-fold cross validation was found to have slightly better (but nearly identical) performance. The kernel SVM (libSVM) was prohibitively computationally intensive, however, as it required more than an order of magnitude more time for classification.

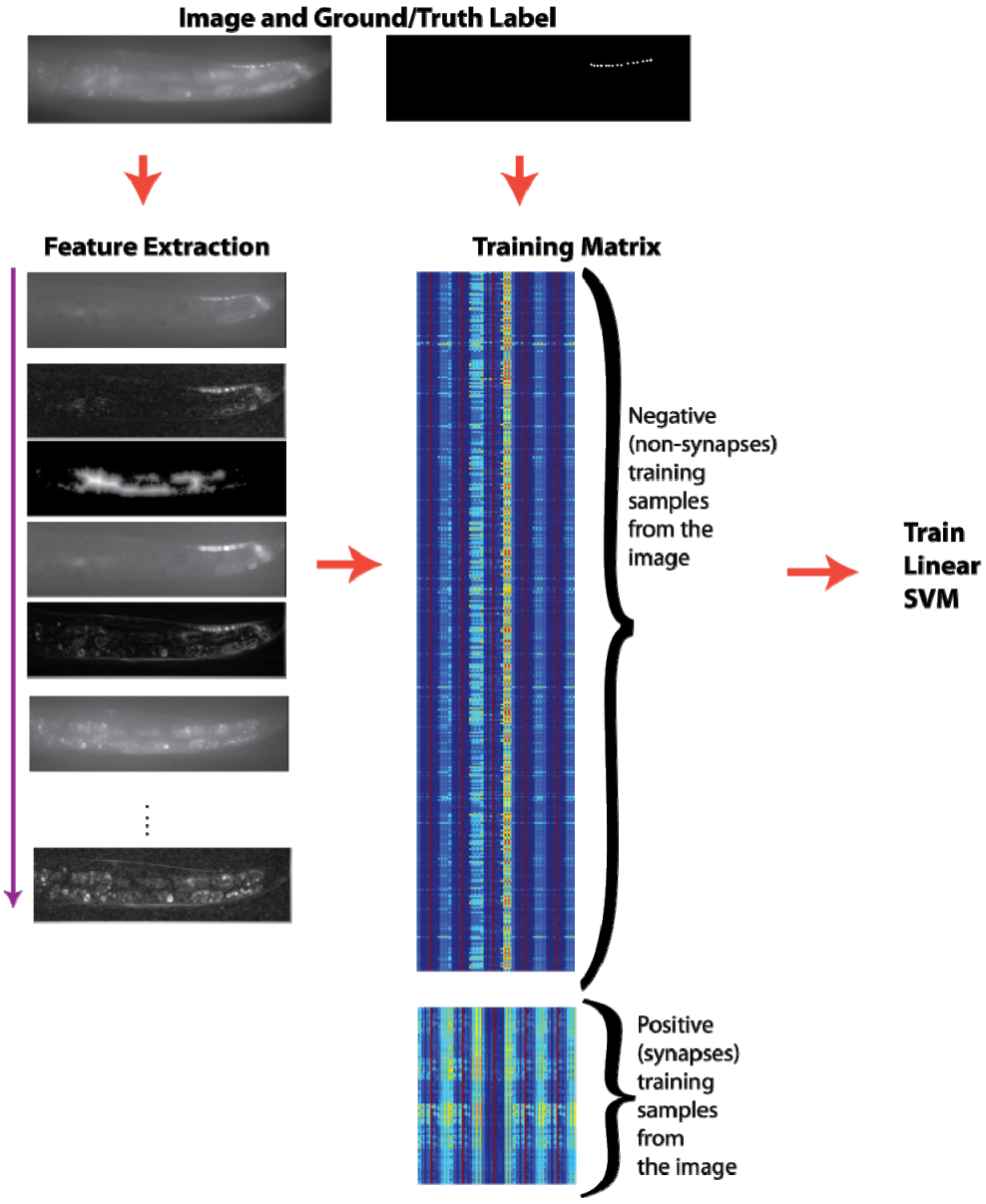



Figure 51: The training of the first stage of the synapse classifier. On each of the ground-truth labeled images, the local features are extracted (feature extraction). Within the images, points are selected for both positive and negative training samples from the ground-truth labels. For each of the pixels selected as positive/negative training, the features are used for training. In the training matrix, the rows are each pixel locations and the columns are the features. Because of the large number of locations that have to be analyzed, a linear-SVM was used for training.

## REGIONAL AND HEURISTIC FEATURES

The second layer of the synapse recognition algorithm serves two purposes.

1) The first purpose is to use the locations of the probable synapses to create extract additional features based on the relative position of the potential synapses to one another.
2) The second is to train a more powerful classifier using just the potential synapse locations as training examples. This classifier has access to the first, local features, in addition to the new features based on the relative locations and image as a whole.

The second layer features use *a priori* knowledge about the biological structure of the synaptic expression and probable mutant expression patterns to extract 30 additional features. The most strenuous biological constraint is that any and all synapses must be formed along the axon or dendrite protruding from the cell body. Of course, there could be significant changes in the guidance cues and thus placement of the axon or dendrite, but this is highly unlikely. Furthermore the development of the specific synaptic pattern is constrained by numerous extra- and intra-cellular cues leading to stereotyped expression along the axon in the wild-type animals. Additionally, synapses are more likely to cluster near one another than to be completely isolated.

There has been considerable work on developing probabilistic models for labeling and segmentation. In particular, conditional random fields (CRF) have been employed to optimize image segmentation by using the relative positions of identified features to more accurately segment images(142-146). In particular, these methods can be used to classify regions that appear accurate based on their immediate surroundings, but less so when viewed in context of the rest of the image. For example, and object classified as a car, which is surrounded by sky and at the top of the image, is more likely to be a plane. A CRF based method was initially implemented but found to be far too slow. The feature extraction for this portion of the method can take at most 1 second, and the CRF based method required tens of seconds.

To optimize for speed, while utilizing the a prior knowledge about the probable synapse locations, a less computationally method was derived. The decision values for the first layer (local features and linear SVM) were used to create an image that expressed the likelihood that any specific pixel is a part of a synapse. Using this decision image as a template for further image processing, many features similar to the first stage were extracted, but within larger neighborhoods. For example, the local maxima and minima of the decision image was extracted using rectangular neighborhoods varying from [5 5] in steps of 5 up to [15 15]. These larger neighborhoods allowed the information on the density of synapses and whether a particular synapse was isolated.

Rather than employing a probabilistic graphical model to learn the relationship between synapse positions, the knowledge that axons tend to be oriented in relatively straight lines was used to incorporate the surrounding pixel information. A Hough transform was applied to a thresholded version of the decision image to identify the two lines most likely to connect the points identified. The Hough transform works by converting groupings of pixels and using a voting method to identify potential lines in the image. The two largest peaks of the Hough transform, corresponding to the most likely lines in the image were identified extracted.

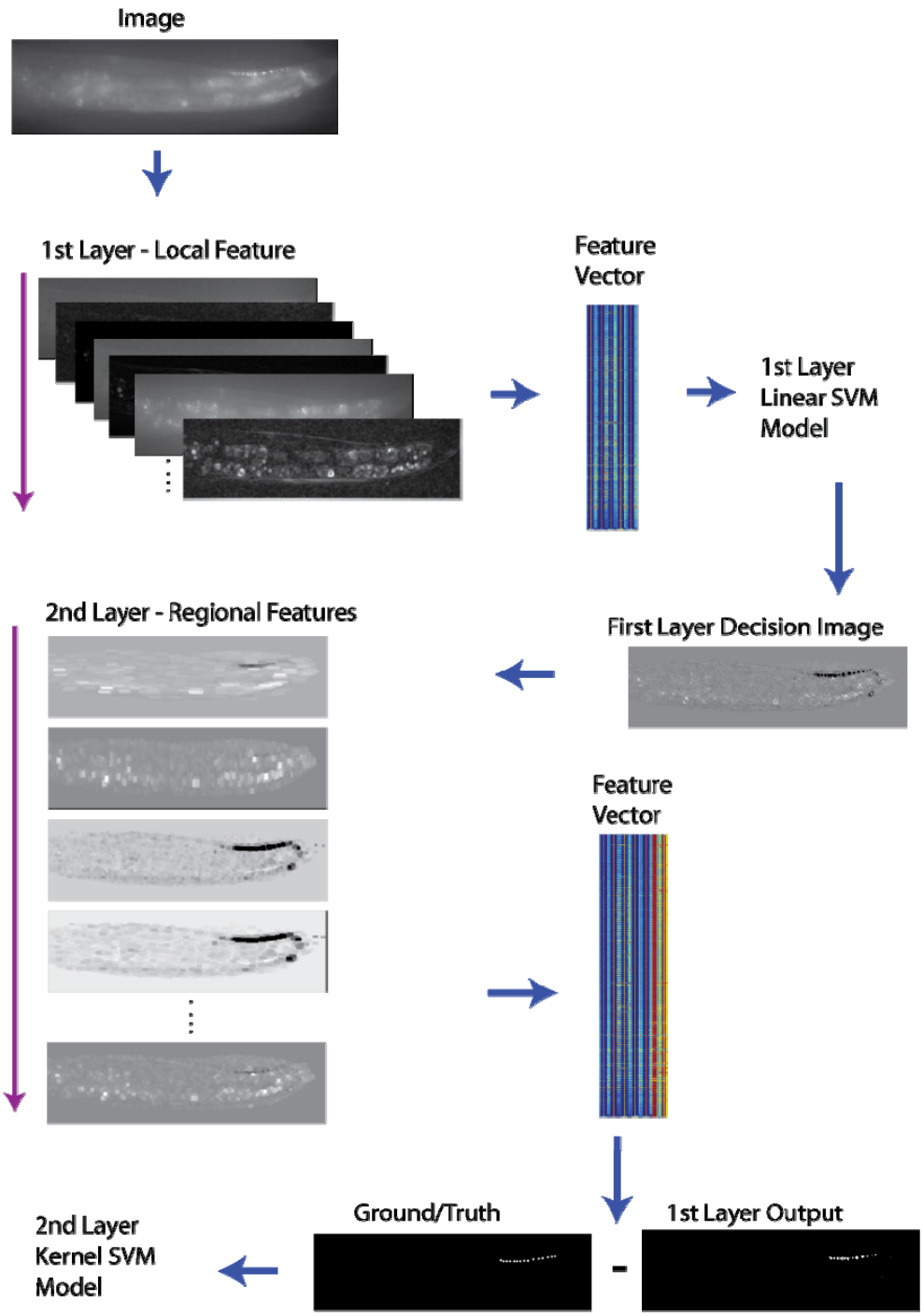



**Figure 52: The training process for the second layer of the synapse classification. The first layer is applied to each of the ground-truth labeled images, and the decision image (containing the probability that each point is a synapse) is used to extract the features for the second layer. The feature vector from the second layer is added to the feature vector from the second layer. Points that were misclassified from the first layer output were used as the negative training class. Training was done using an RBF kernel-SVM.**

Using these two directions, several morphological features were applied to the image with the raw decision values. These included erosion and dilation using differently sized structuring elements and performing a two-dimensional convolution using lines of different lengths, in the directions of the two primary lines in the image. The directions perpendicular to those identified using the Hough transform were used in addition to primary. If a sequence of low probability synapses were aligned close to one another, these operations would extract that information and allow the second stage classifier to use that information. Likewise, a single point that is far from any other potential synapses is more likely to be an incorrect classification, and these features will reflect that.

For each of the locations, the feature vector from the second layer features is combined with the feature vector from the first layer. This creates a feature vector that contains 145 features for each pixel location. Because the linear SVM from the first layer was trained to err on the side of misidentifying non-synapses as synapses, the second layer only looks at locations identified by the first as being potential synapses. This serves several purposes. First of all, it reduces the number of data points dramatically (by 3 orders of magnitude) and thus decreases the computational time required for classification accordingly. Secondly, this decrease in computational time increases allows the use of a more robust, but slower classification using an RBF-kernel SVM. Thirdly, it allows the classifier to be trained only on points that were misclassified in the first layer and thus provide an increase in classification accuracy. This training process is shown in Figure 52.

Using the ground-truth library of images, this second stage incorporating the regional features was trained. A brute force grid search optimization method using 5-fold cross-validation was employed to determine the optimal training parameters. The optimal parameters identified during the grid search were $C = 2\ and\ \gamma = .0625$.

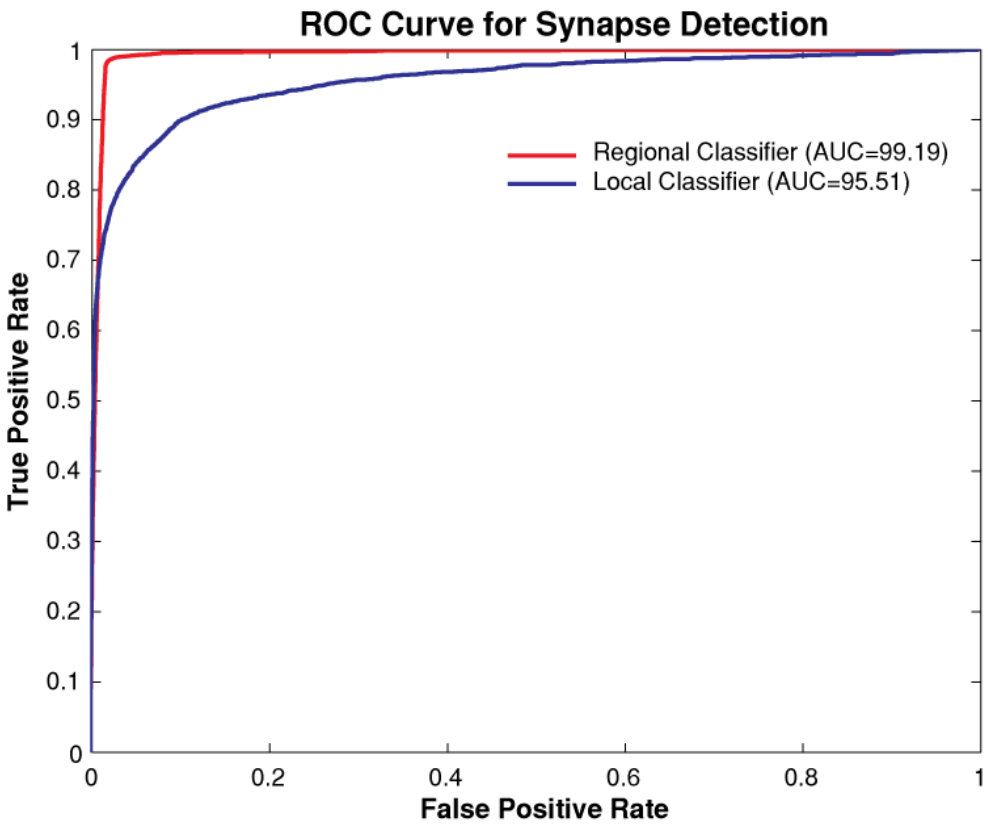



**Figure 53: ROC curves for the two synapse classifiers calculated using the pixel-by-pixel method on the ground-truth labeled library. The error rates were determined using 5-fold cross-validation and a grid-search for parameter optimization for each classifier. Because it is a rare-search problem, the 3,000 points in each image most likely to be synapses were used to determine false and true-positive rates.**

In order to determine the performance of the synapse classifiers, pixel performance statistics were calculated for each of the two methods. Performance characteristics where determined using each pixel within the image, and whether the individual pixel should be a synapse or not. The library of ground-truth labeled images was used for training both of the classifiers. Because the two-stage classifier incorporating the regional features is dependent on the first layer (local features), the image library of images was divided into training and testing sets for 5-fold cross validation. Both classifiers were trained on the same image set and tested on the same image set. During each of the cross-validation performance tests a grid-search was performed to determine optimal training parameters for the training. Because identifying synapses is a rare-search problem, the vast majority of pixels were correctly rejected using both

classifiers. Using all of these data points obscures the performance differences between the two methods, so for each image the 3,000 points most likely to be synapses were used to determine classifier performance. Employing this method, the ROC curve in Figure 53 was calculated. As can be clearly seen in the figure, the two-stage classifier that incorporates regional features performs significantly better than the classifier only using the local features. Individual measurements of classifier performance derived from these curves can be seen in Table 10.

**Table 10: Statistics comparing the performance of the synapse classifiers using pixel-by-pixel performance. The two-stage classifier incorporating the regional features has significantly better performance relative to the classifier only using the local features. These statistics were derived from the ROC curves above. AUC is the area under the curve, EER is the equal error rate (when 1-TP = FP), and the True Positive rate when False Positives equal 1% and 0.1%.**

| | AUC | EER | TP @ FP = 1.0% | TP @ FP = 0.1% |
|---|---|---|---|---|
| Local Features | 95.51 | 10.12 | 69.00 | 22.50 |
| Regional Features | 99.19 | 1.87 | 74.37 | 14.98 |

## ALTERNATIVE REGION-BASED FEATURES

The two-stage classification system was what was ultimately implemented for successful synapse identification and automated screening. Leading up to the design of this system, however, numerous alternatives were implemented and evaluated. These methods proved insufficient for reasons ranging from requiring too much computational time, to resulting in a high miss-classification rate. The goal in using these region based features was to either identify probable regions where synapses would be located, and then only search within those areas using a single pixel classifier, or to combine the decision values from the two classifiers and train a third classifier to use these two inputs to identify synapses. These two approaches are shown in Figure 54.

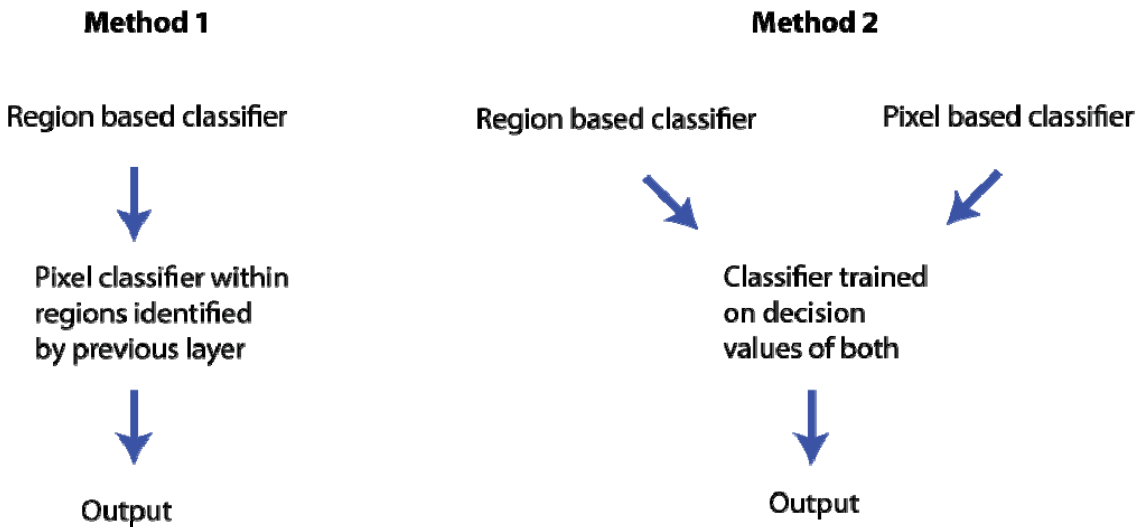



**Figure 54: The two methods of combining regional based classifiers and pixel level classifiers. Both methods were attempted for the region based classifiers discussed in this section, but ultimately the classification error didn't justify the computational time required for these approaches.**

The method used was the Histogram of Gradient (HOG) features as detailed by Navneet Dalal(101). These features were originally designed to identify people in images and video by binning the angle and magnitude of the gradient within small neighborhoods, called cells. These cells would then be grouped into overlapping groupings of several cells called blocks. By densely overlapping these cells, the blocks could incorporate some of the features from the surrounding region to gain an understanding of the image on a larger scale. The HOG features, however, proved to perform fairly poorly when working with the fluorescent images in this project. The majority of the fluorescent objects in the image are circular in contrast with images of everyday, where the direction of the gradient offers a significant amount of information about the object in question. Because of this, the basic implementation had significant difficulty in distinguishing between fat granules and synapses. Additional features were added to each of the cells, including histograms of the difference of Gaussians and Laplacian of the Gaussian. These additional features improved the performance, but not to a point that would warrant the computational time required. Using the HOG features, the methods shown in Figure 54 were applied to train classifiers to identify the locations of synapses. The first method proved faster as the local features were only

extracted on a very small area, but was prone to errors. The second method proved more successful, and resulted in a reduced error rate.

## PHENOTYPICAL FEATURES FOR ANIMAL IDENTIFICATION

Following the identification of the synapses, it was necessary to extract specific features that could be quantified and used for phenotyping purposes. These features were modeled after features that had been used successfully by the Shen lab to identify genes affecting extra-cellular guidance cues during DA9 synaptogenesis, as well as additional phentoypical quantifications that could be of potential interest. These additional features were extremely similar to those we used in the preliminary tests of automated screening(66) discussed in Chapter 4.

Many of the genes involved in regulating the formation of synapses in DA9 are used to determine the specific spatial locations that where synapses should form, and where formation is inhibited. The first group of features is related to the relative positions of the synapses, both to each other and to landmarks external to the synapses. These features provide information about the signals guiding synapses formation. The second set of features is based on things that provide information about the individual synapses

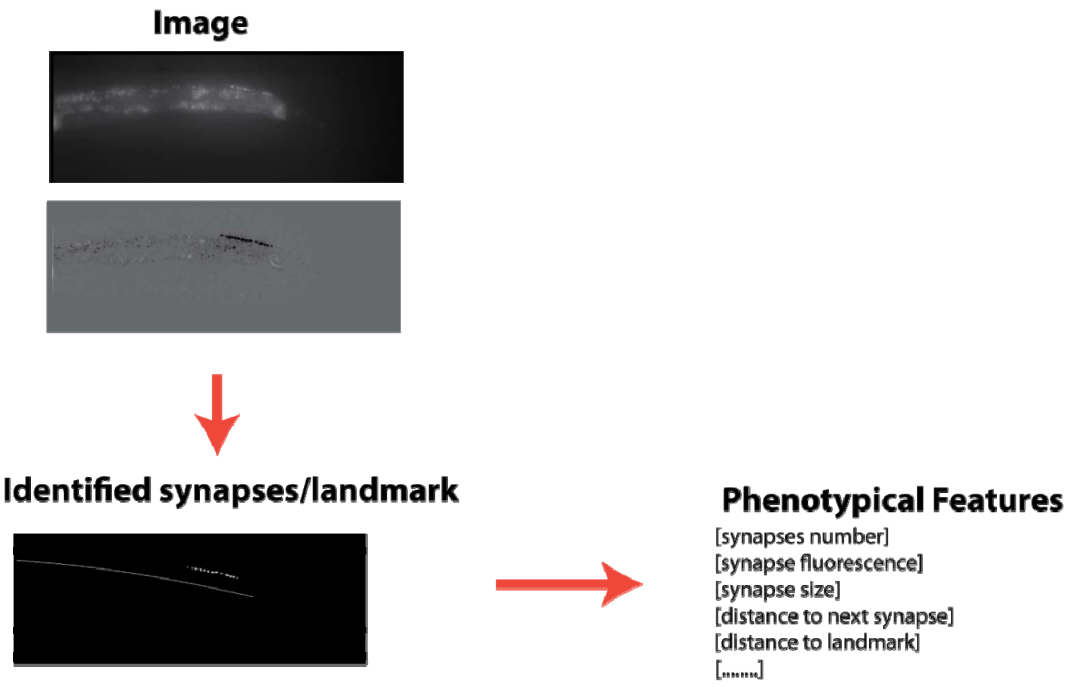



**Figure 55: The phenotyping process. After the animal is imaged and the synapses are identified and extracted, landmarks are identified. In this work, the landmarks were the midline and the end of the intestine. After this, specific quantitative features are extracted that describe both the individual synapses (size, intensity, etc) and the distribution of the synapses relative to one another and in relationship to the biological landmarks.**

## LANDMARK IDENTIFICATION

Once the synapses have been located and identified, it becomes important to identify other features, external to the synapses themselves that can be used for extracting quantitative features. In some cases, this could be the distance of synapses to a specific cell, or the number of synapses that are clustered into specific areas on the animal. To do these, however, the landmarks must be identified. This would be easiest served by using a fluorescent reporter that could label the landmark of interest. Ideally, this fluorescent reporter used would be in a different spectrum than the synapse reporter (yellow or red in this case) to make it easier to identify the landmark.

In this case, however, because no fluorescent reporters were present in the area nearby the synapses, several alternative landmarks were identified and used. The landmarks identified for this project were the

midline of the animal and the end of the intestine. Because synapses are only supposed to be located on the ventral side of the animal, correctly determining whether all the synapses are on the same side is important, and is done using the animal midline. This can be done relatively easily because the autofluorescence of the animal is significantly greater than the surrounding areas. The midline is determined by multiplying the red and green channel images, and thresholding for any locations greater than the mean. This image was then thinned to identify the midline. Using both spectral images improved the SNR and reduced the odds of misidentifying the midline.

The second landmark used was the end of the intestine. Whereas the first landmark serves to identify whether individual synapses are located on the ventral or dorsal side of the body, the second feature helps to whether synapses are in the asynaptic regions posterior to the DA9 cell body. The stereotyped position of the intestine and the relatively strong levels of fluorescence make the intestine an appealing landmark. The end of the intestine corresponds very similarly to the location of the beginning of the asynaptic zone in wild-type animals.

## PHENOTYPICAL FEATURES

The features extracted for the purpose of this project were selected both in reference to previously published work, *a priori* biological knowledge, and features invented merely because they seemed like they might be relevant. This is the particular advantage of the machine learning approach to biological phenotyping – the abundance of irrelevant features is not necessarily harmful to the work. This system has been transferred to our biological collaborators (Shen Lab at Stanford). Development of new phenotypical features that quantify the specific biological phenomena of interest to them is currently underway.

For each of the synapses located in the image, the location relative to the two landmarks is identified. Then, features concerning the synapses and its relationship to the other identified synapse are extracted.

These include features such as the distance to the nearest and furthest synapse. Additionally, the number of synapses located within a small radius around the synapses is extracted. Following these, features regarding the individual synapse in isolation are extracted. These include features such as the size, ellipticity, intensity, and major/minor axis length.

# CHAPTER 6: AUTOMATED MUTANT DISCOVERY BY AUTOMATED SCREENING

This chapter details the application of the completed microsystem to perform large-scale automated screens based on subtle changes in synapse patterns. Having created a system to identify synapses within an image and extract, specific phenotypical features, the next step is to demonstrate the utility of the system to a real-life problem. The combined microfluidic device and computer vision framework would be capable for either reverse or forward genetic screening, but after discussion with our collaborators (Shen Lab, Stanford), we chose to perform a forward genetic screen. A forward screen was selected as it offers the opportunity of discovering novel alleles of previously identified genes. Furthermore, although RNAi works quite easily in *C. elegans*, the knockdown of gene function can vary dramatically especially in neuronal screens. This increases the likelihood that some genes that affect the DA9 synaptogensis network could be overlooked. Because of the difficulty of performing these screens manually, the screens in this chapter would have taken a postdoctoral student over a year to complete.

Performing a forward genetic screen requires several steps prior to running the actual screen and extracting novel mutants

1) The wild-type population containing the fluorescent reporter of interest must be phenotyped. This requires image a sufficient number of animals under the same conditions as the eventual screen, and then using those images to extract the phenotypical features of interest. This provides a point of comparison for any mutants.
2) The method of screening and classifying animals as either mutants of interest or wild-type must be determined. The two choices are: a trained classifier using the wild-type and a known mutant for training, or an outlier detection screen that looks for anything that seems different from the wild-type population. If performing a screen using a classifier trained against a known mutant,

sufficient images for phenotyping the known mutant must be acquired. Both methods were used in this work.

3) The mutagenesis method and methodology must be selected and performed. For this work, standard EMS protocols were followed.
4) Following screening, the animals identified must be cloned, decontaminated and then verified. This is done visually to verify animals of interest, and then animals are sequenced. Due to the time required for sequencing, only a few of the new mutants have been sequenced to date.

## ESTABLISHING THE PHENOSPACE WITH WILD-TYPE AND KNOWN MUTANTS

In order to perform a screen to identify novel mutants, it is necessary to have an idea of the phenotypical space. In typical, manual, screens this is also performed. In these cases, however, creating the phenotypical space, and understanding the wild-type distribution is a qualitative training process. A biologist would never screen for new mutants without understanding what the wild-type population looks like, or without having an idea of what phenotypes might be of interest. The process by which a biologist becomes familiar with a new reporter, or expression pattern, is identical to my process of establishing the quantitative phenospace using the computer vision system described previously. Having a robust understanding of the phenotypical space of the wild-type population, and what features are important allows accurate decision making between wild-type and potentially novel mutants. Just as when training a human decision maker, a certain number of animals must be imaged and phenotyped to allow a robust understanding of how the wild-type population is modeled. In order to ensure that the phenotyping conditions were as close as possible to the screening conditions, all of the imaging and phenotyping was done using the microfluidic system.

## ANIMAL GENOTYPES AND DETAILS

The animals imaged and used in these experiments all contain the transgene *wyIs85* (pITR-1:GFP::rab-3). This is a synaptic localized GFP reporter that appears exclusively in the motor neuron DA9. The itr-1 promoter is turned on at the early L4 stage. Because the Shen lab has used the DA9 model system for several years to study genes affecting synaptogenesis, they had previously identified several genes that expressed altered synaptic patterns. These animals with altered expression patterns allowed me to image a population, and train a classifier to distinguish between them and the wild-type population. The Shen lab provided several mutants with altered expression(18, 21, 147). This situation, of having access to multiple mutant phenotypes, is a relatively uncommon situation. For most of the screens that would be interesting for screening and additional work, researchers would have access to perhaps a single, or no, mutants before beginning screening. For this reason, rather than using all of these animals to populate the phenospace and train the classifier, as it would be uncommon, a single mutant genotype (*lin-44*) was selected for imaging and training. The genotypes of the animals used in my work are as follows:

XA7810 (*wyIs85*) and XA7812 (*wyIs85* ; *lin-44* (n1792))

As discussed in the previous chapters, the wild-type animals express synapses in tightly controlled locations along the dorsal axon. The mutant selected has altered expression with synapses present in the asynaptic zone of the dorsal axon, and occasionally synapses in the commissural axon.

For phenotyping purposes hundreds of z-stacks were acquired of both the wild-type and mutant animals. Ultimately, 320 wild-type (XA7810) animals and 97 mutant (XA7812) animals were used for phenotyping. Because it was important to acquire images that accurately represented the actual screening conditions, the images were acquired in an automated fashion without any human intervention. This resulted in some of these images being blurry or having synapses that were distorted due to reflections or the cuticle acting as a lens. The acquisition of non-ideal images that conformed closely to the actual images acquired

during screening was important as it allowed the phenotyping to capture the range of phenotypical expression that would be extracted in a screen.

Following imaging, the animals were processed using the computer vision framework from the previous chapter to identify synapses and extract the phenotypical features. The results of the imaging and phenotyping of XA7810 and XA7812 can be seen in Figure 56. Because the phenospace is many dimentional, in order to display the extracted phenotypes, the data was compressed using the principle components. As can be clearly seen in the image, there is significant separation between the two genotypes, even in this low-dimensional space. The variation and heterogeneity in the populations is a function of several factors. Within each image, depending on the image quality and the synapse resolution, varying numbers of synapses will be identified. Additionally, when identifying the landmarks used in phenotyping, the biological features have a small amount of variation, and the algorithm introduces significantly more variation. The combination of these factors results in the broad distribution seen in the Figure 56.

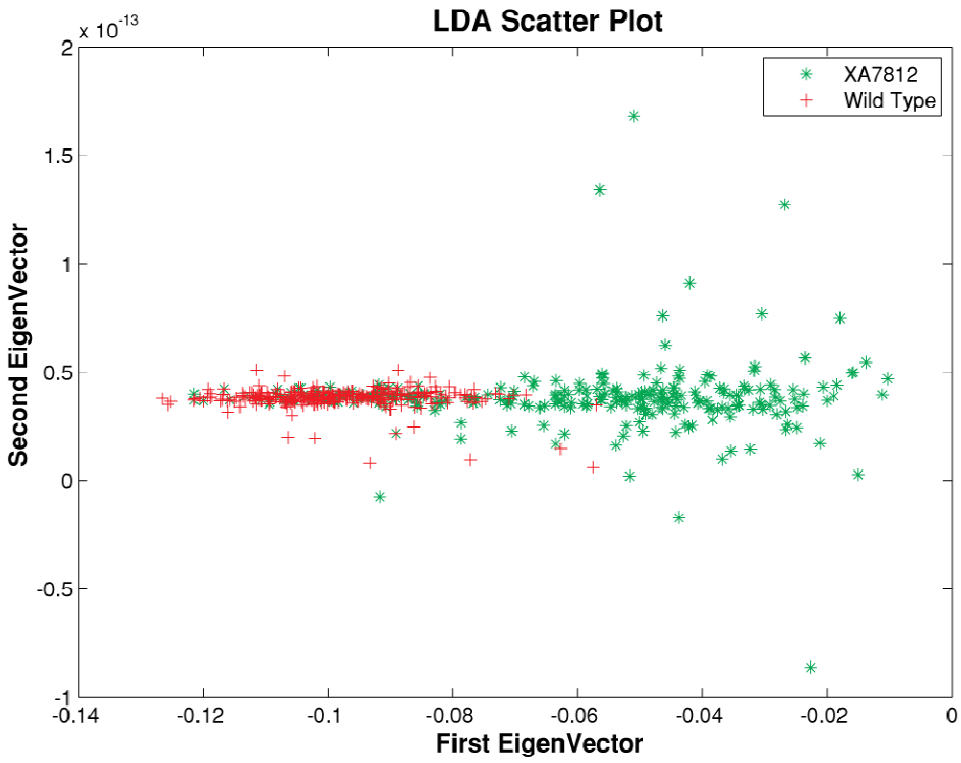



**Figure 56: The compressed phenospace of the wild-type population and one of the previously discovered mutants. To reduce the dimensions of the phenospace, the values were normalized on a 0 to 1 scale, and then the eigenvectors were computed. The two populations can be relatively easily separated on this plot, which means that both methods of screening would likely identify this or similar mutants.**

## SCREEN METHODOLOGY

The screens were all performed according to standard protocols for forward-screening. There are several mutagen options for forward screens including UV radiation, ethyl methanesulfonate (EMS) among others. Based on the ease and consistency of the results, the EMS method was selected. EMS is potentially carcinogenic chemical that induces point mutations in the genome. It works by changing G:C base pairs to A:T. The dosing and incubation time produce reliable numbers of mutations, which makes it a convenient tool as the number of haploid genomes that have been screened can be easily estimated from this process.

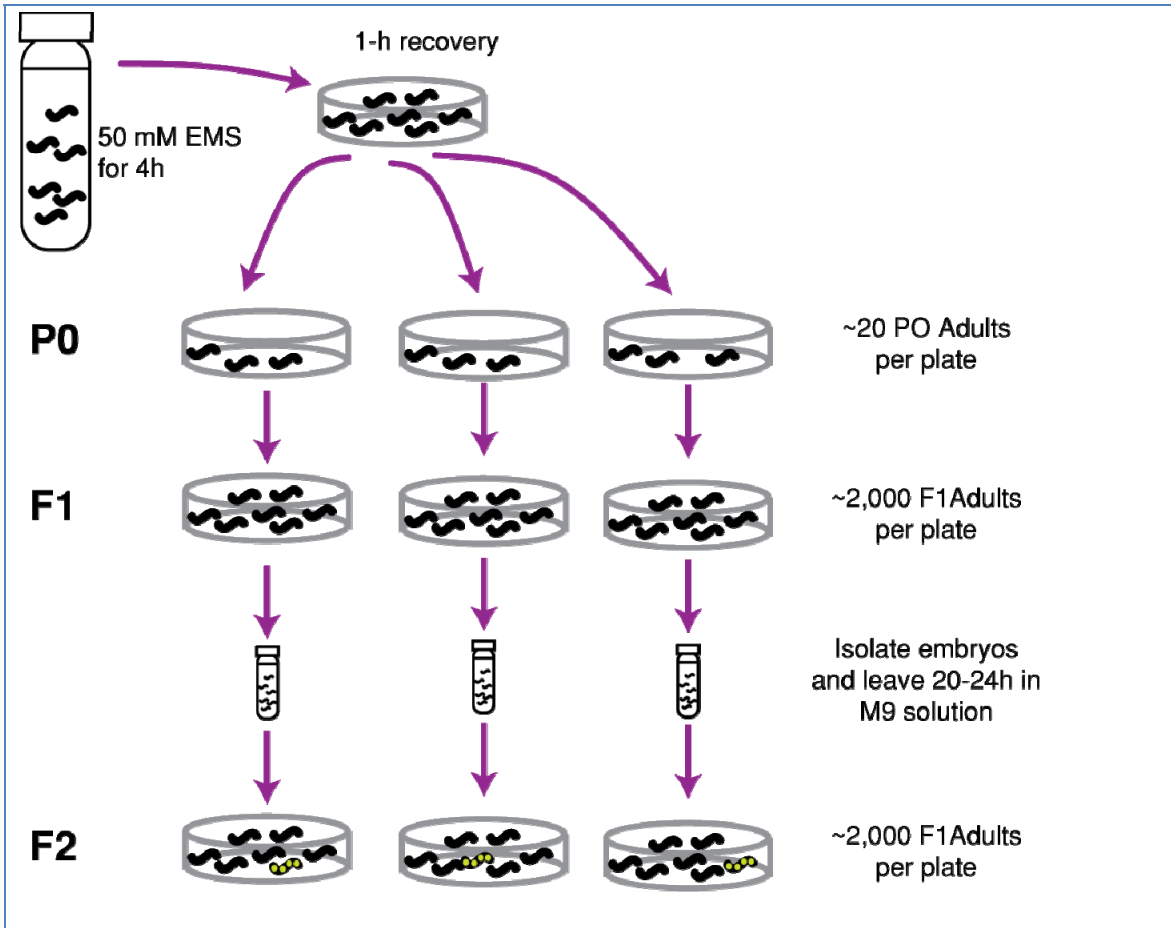



**Figure 57: The mutagenesis protocol used in the screens. Synchronized L4 animals were exposed to the EMS for 4 hours, and then placed onto individual plates. Once the F1 animals reached adulthood, they were bleached and embryos isolated. To age-synchronize the embryos, they were left in M9 buffer for 20-24 hours and then grown on NGM plates until young adulthood.**

For this project, the EMS protocol was performed as is shown in Figure 57. All of the steps involving EMS were performed in a chemical hood, and all fluid or materials that were contaminated with EMS were rinsed or diluted with 1M NaOH. To create the mutagenized animals for screening, a population of adult animals was bleached to isolate the embryos and age-synchronize the offspring. When the animals reached the L4 stage, a large number of synchronized L4 wild-type animals (containing the *wyIs85* transgene) were suspended in a buffer solution and 20mM EMS for a period of four hours. Following the four hour incubation period, the animals were then rinsed with the M9 buffer solution and placed on plates. These P0 animals are allowed to grow to adults, and once the F1 progeny from these animals become adults, the F2 offspring are age synchronized. Because the microfluidic device is sensitive to

dramatic changes in the size of the animals, and mutagenesis creates a rather significant variation in growth rates, additional steps were taken to age-synchronize the animals. Once they have reached the proper developmental stage, they will be screened in the device. F2 eggs were obtained by bleaching F1 adults using a solution containing about 1% NaOCl and 0.1 M NaOH. This isolated the adult embryos. The embryos were then washed in M9 buffer and grown for 24 hours. Because the animals do not have access to food, they halt at the L1 stage after hatching. Leaving them overnight ensures that animals are relatively tightly synchronized at the L1 stage when they are placed onto plates containing Nematode Growth Medium (NGM) seeded with *E. coli* OP50.

## OVERVIEW OF SCREENING OPTIONS

During screening, the general problem is to determine whether an animal appears to have a wild-type expression pattern, or could be classified as having a novel expression. This classification process can be approached in two different ways: trained detection and outlier detection. Fundamentally, the objective of both approaches is identical, but the different implementations lead to distinct advantages and disadvantages for each method.

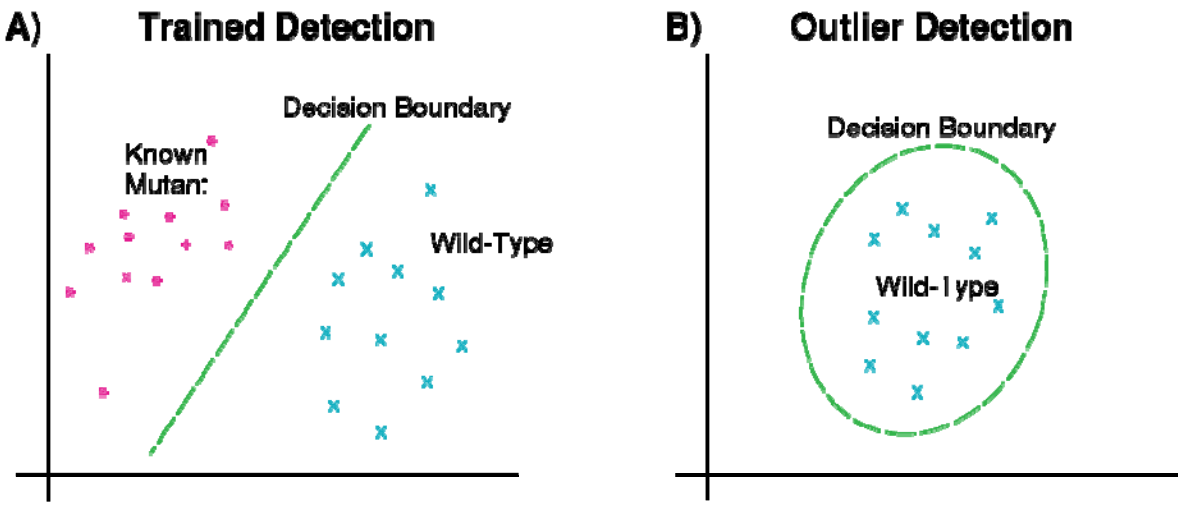



**Figure 58: The potential screening methods. (A) Trained detection using a known mutant and wild-type**

A screen employing a trained classifier uses a classifier that has been trained on both the wild-type population, and a mutant with an altered expression pattern. This is a discriminative classification method as it uses two populations, and is trained to identify the differences between them. In the trained classifier case both populations of animals are imaged and phenotypical information extracted. Using this information, a classifier can be trained using machine learning methods that can distinguish between the wild-type and mutant populations. This method is shown in Figure 58A.

The alternative to this is outlier detection, which is a generative statistical method. Rather than training a classifier to distinguish between two populations, outlier detection attempts to model the underlying distribution of the wild-type population in the phenospace. If this wild-type distribution can be accurately modeled using the phenotypical features extracted, it is possible to look for animals that would lie at the edge of the distribution, or have a low probability of arising from the wild-type distribution. A schematic showing how this would work is shown in Figure 58B. Rather than classifying animals as either population, the probability that the animal would naturally occur in the wild-type population is estimated. Then, if the percentage is below a specific user determined cutoff (typically ~1%), the animal is classified as a mutant.

The use of a discriminative classifier brings specific advantages and disadvantages that must be evaluated prior to use. Using a discriminative, or trained, classifier is, of course, predicated by the ability to use a known mutant for training purposes. If there are no known mutants with altered expression patterns, then this method cannot be used. Using only a single mutant with altered expression for training could cause the decision boundary to be drawn in such a way that new mutants could be classified as appearing more similar to wild-type than to the previously discovered mutant. Additionally, because features that have limited discriminative power between the two populations are not used for classification, new classes of mutants that have altered expression in features that were discarded during training could be overlooked. Because features that are irrelevant are discarded during training, however, it is possible to invent as many features and measurements as possible. Some of these features will be useless, and many

will be only marginally helpful, but the combination of features weighted and used. At the worst, additional features will leave performance unchanged, but unlike outlier detection, will not decrease performance. This means that no *a priori* knowledge is needed to create the feature set, and thus a unified set of features could be used for all problems as irrelevant features would be removed during training.

While some of these tradeoffs for a discriminative classifier can be compensated for with a generative classifier, it also has inherent advantages and disadvantages. The disadvantages of outlier detection are primarily that it requires more *a priori* knowledge regarding the selection and use of the phenotypical features. Because the wild-type population is modeled in the phenospace, the feature selection is critical. The selection of a large number of features that are irrelevant and of limited ability to detect new mutants can degrade the detection ability. Of course, the other side of this is that creative development of features can be advantageous because it allows the user to look for novel classes of animals that are different than anything previously detected.

**Table 11: A summary of the advantages and disadvantages of both classification methods used during mutant screens.**

| | **Advantages** | **Disadvantages** |
|---|---|---|
| **Trained Classifier (Discriminative Classifier)** | • Focuses on only the important features from the phenospace. Marginal or irrelevant features will not decrease performance.<br>• Requires no knowledge of relevant phenotypical features. | • Requires at least one genotype with expression different than wild-type<br>• Biased towards features that differentiate between wild-type and training set.<br>• New classes of mutants could appear more similar to wild-type and thus be undetected. |
| **Outlier Detection (Generative Classifier)** | • Does not require previously discovered mutants<br>• Provides flexibility to choose features of interest and identify novel classes of animals | • Requires *a priori* knowledge for the feature creation<br>• Marginal or irrelevant features can decrease performance. |

Because each method has specific pros and cons, the system was created to allow for screening using either method. This would allow it to be generalized both to problems where numerous mutant

genotypes with altered expressions are available, and problems where it is of interest to discover completely novel classes of animals. Two large-scale screens were performed, one using each of these methods. This demonstrates the power and generalizability of this framework.

## AUTOMATED SCREENING USING TRAINED CLASSIFIER

The screen using a trained classifier was performed using the wild-type (XA7810) and a *lin-44* (XA7812) background strain for training purposes. In most screening situations, researchers would have access to a single, or only a few animals with altered expression profile. Because of this, I chose to only train the classifier against a single mutant strain. This demonstrates the effectiveness of the screening method in the most generalizable condition. The use of additional mutant genotypes for training purposes would only improve the throughput. The features from the wild-type and mutant populations were normalized on a [-1 1] scale and then trained using an RBF-kernel SVM. The cost and gamma for training were determined using a five-fold cross validation to minimize the mis-classification error. Due to the relatively small feature vector length, the classification time is <100 ms.

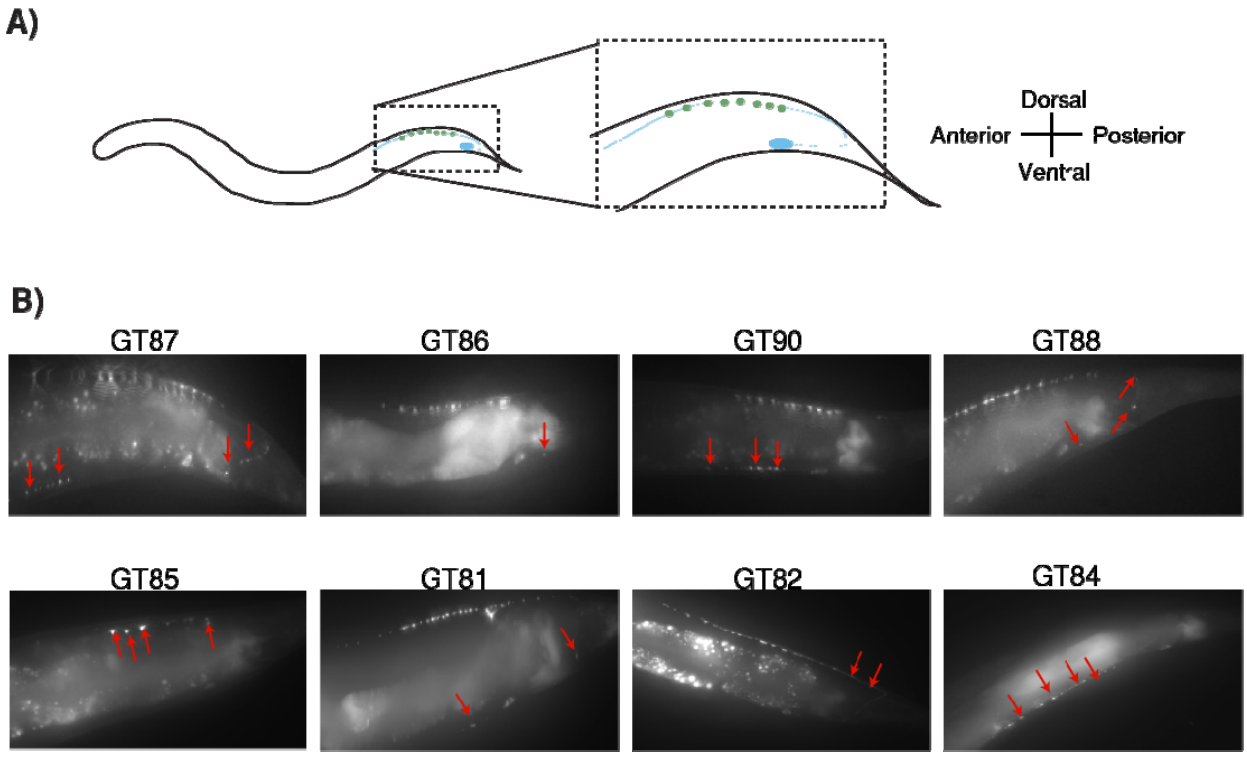



**Figure 59: Selected images from the trained classifier screen. (A) Schematic showing the portion of the animal displayed in the images below. (B) Representative images from each of the genotypes displayed. The red arrows in the images point to synapses that are at an altered location or have an altered morphology.**

Validation of the identified mutants was performed in several stages. Initially, animals were imaged manually and verified that that they did have an altered expression pattern compared to wild-type. Examples of the animals discovered in the discriminative screen can be seen in Figure 59. All the animals in the figure are oriented identically, and the red arrows point to specific features that distinguish them from the wild-type population. Although the visual validation is sufficient to determine whether animals are actual mutants, or merely false-positives, additional work needs to be done to identify whether they are new alleles of genes previously identified as components of the specific pathways, or the discovery of genes previously unknown as part of the pathway. The results from the screening including the throughput and screening results are shown in Table 12.

**Table 12: The results from the discriminative classifier (trained) screen.**

| Animals Screened | Average Throughput | Sorted as mutants | Verified Mutants | Novel Classes of animals | New Alleles of known genes | Awaiting classification |
|---|---|---|---|---|---|---|
| 2100 | 191/hr | 32 | 10 | 1 | TBD | 9 |

## AUTOMATED SCREENING USING OUTLIER DETECTION

The screen using outlier detection was performed following the screen using the discriminative was performed using the same collection of wild-type (XA7810) images that was used to train the discriminative classifier. All features dimensions were normalized to a [-1 1] scale to ensure that a single feature didn't dominate. The population of wild-type animals was modeled in the phenospace as a single multi-dimensional Gaussian distribution. The cumulative distribution function for each new animal was then used as the cutoff threshold for classifying animals as wild-type or mutant. The CDF from the library of wild-type animals was used to determine the mean CDF for the wild-type population. A threshold two-orders of magnitude lower than that value was used during this screening process. This could be easily increased or decreased depending on the tolerance for false positives.

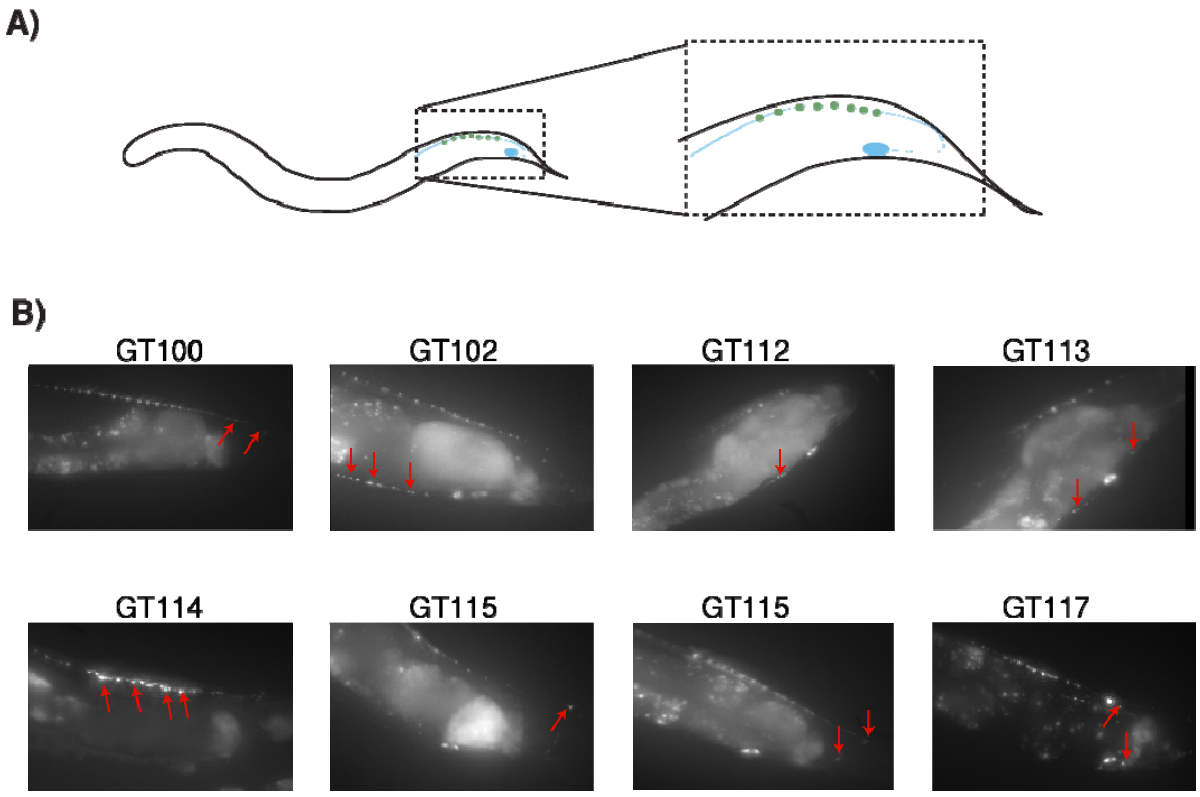



**Figure 60: Images of some of the animals identified during the outlier detection screen. (A) Schematic showing the portion of the animal displayed in the images below. (B) Representative images from each of the genotypes displayed. The red arrows in the images point to synapses that are at an altered location or have an altered morphology.**

Following screening, the extracted animals were verified in the same method as with the previous screen. Examples of the animals discovered in the outlier detection screen can be seen in Figure 60. All the animals in the figure are oriented identically, and the red arrows point to specific features that distinguish them from the wild-type population. The throughput was slightly higher in the outlier detection screen, but this was primarily due to a few tweaks in the error handling and processing based on observations made during the discriminative screen. The computational processing time for each method is equivalent, and as such, no conclusions should be drawn that one method was faster than the other.

**Table 13: The results from the outlier detection screen.**

| Animals Screened | Average Throughput | Sorted as mutants | Verified Mutants | Novel Classes of animals | New Alleles of known genes | Awaiting classification |
|---|---|---|---|---|---|---|
| 3100 | 243/hr | 52 | 18 | 2 | TBD | 16 |

## BIOLOGICAL VALIDATION OF NEW MUTANTS

Once an animal was identified as a potential or probable mutant during the screening, they had to be cloned and then verified. Because *C. elegans* are hermaprhodites and self-fertilizing, the cloning process is straightforward relative to other multicellular model organisms. Following screening, all of the potential mutants were picked onto individual NGM plates. Animals are left on the plates for several days to reproduce. In addition the mutations of interest that caused the changes to the synaptic reporter expression pattern, there are additional background mutations present in each of the animal (typically ~5 additional point mutations). Not only are animals selected as a result of changes in their synaptic expression, but this selection process coupled with these additional background mutations causes many of the animals to be extremely sickly. Not only does this result in additional generation time, but typically ~30% of the extracted animals are sterile. Additionally, as a result of animal handling, extracted animals are often contaminated with a mold, fungus or bacteria that require decontamination.

Once the identified mutants have been cloned and decontaminated, it is necessary to verify that the animals are probable mutants. Misclassification of fat granules as synapses, and vice-versa, will result in animals being misclassified as mutants or wild-type. Because it is impossible to determine the false negative rate (mutants classified as wild-type), only the false positives are identified and reported. Furthermore, because genetic validation by sequencing is such a slow and time consuming process, it is important to visually confirm that each of the animals to be sequenced have an altered synaptic expression. Validation was done using standard worm protocols using a worm-slide and sodium azide for immobilization. A small population (10-15) animals of each genotype was imaged, and these images were used to determine the accuracy of the screening. Each of the genotypes classified as mutants of interest were named and are in the process of being sequenced. The results of the sorting are shown in Table 14.

For the purpose of these results, animals that were sorted as mutants but failed to produce any offspring that could be used to verify whether the animals were correctly or incorrectly sorted, were removed from the results. False positives were calculated using the following equation where animals sterile animals were removed from the denominator: $FP = \frac{non-mutants}{Animals\ sorted\ as\ mutants}$. The upper bound of the enrichment was calculated using the equation: $Enrichment = \frac{Mutants/_{Sorted\ as\ mutants}}{Mutants/_{Animals\ processed}}$. Because the number of mutant animals that were incorrectly sorted as wild-type cannot be verified, they were assumed to be zero for enrichment purposes.

**Table 14: The results of automated screening for DA9 synaptic mutants.**

| | Percentage sorted as mutants | Potential Mutants | Visually Verified | False Positive Percentage | Enrichment (upper bound) |
|---|---|---|---|---|---|
| **Trained Classifier** | 1.5% | 32 | 10 | 69% | 6600% |
| **Outlier Detection** | 1.7% | 52 | 18 | 65% | 5900% |

These identified mutants contain numerous classes with altered reporter morphology. In order to more accurately visualize some of the expression patterns within the quantitative phenotypical space, several populations of animals from the discriminative classifier screen were re-imaged in the device. These populations of animals and the imaging results could be used in the future for improving the discriminative classifier by including these populations in the training with wild-type *vs.* all known mutant training. Because of the range of phenotypes present in this new training set, it would increase the number of features classified as important, and more accurately reflect the features that could be used to identify novel mutants. To visualize this selection of animals on the quantitative phenospace, the data was projected onto the two principle components of the wild-type animal features. The results of this projection can be seen in Figure 61. The distributions of all the mutant animals are clearly distinct from the wild-type population. Due to the noise in the phenotypical features and issues regarding penetrance, some overlap between mutants and wild-type would be expected. In addition to perform screening of

animal populations to identify new and novel mutants, there is the possibility of phenotyping animals more precisely than previously imagined, and extracting information such as penetrance or predicting the potential gene mutation. Figure 61 provides some support that this could be performed using my system, but additional work would have to be performed to make this a reality.

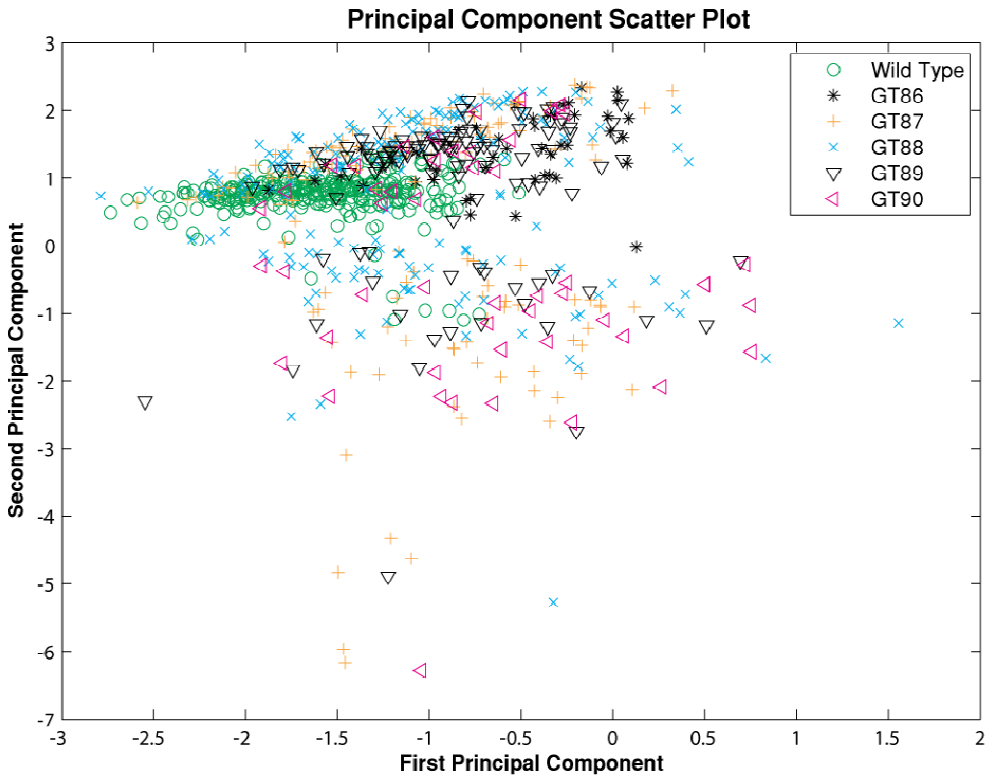



**Figure 61: The phenotypical information for several of the animals identified during the discriminative classifier screen. These animals were re-imaged using the microsystem, and could be used for improving the discriminative classifier. All features were projected onto the principle components of the wild-type population.**

This mutants discovered during the screens are currently being analyzed. Initially, this is done visually to identify potential mutations that could cause similar synaptic patterns. If there are any potential targets, a complementarity test with these genes known to produce similar expression patterns is perfromed. If the new animals are not complementary to any previously identified genes, the animals are sequenced to identify the new genes. Because this is a slow process, sequencing of the newly discovered animals is

currently underway and only some of the animals have had the complementarity test performed. In Table 15 the animals from the discriminative screen are shown. Only animals that were verified to possess altered synaptic expression patterns were included in the tables. Of particular interest are the animals two animals that are a novel class of animals. This demonstrates the power of the microsystem to both screen for new genes involved in previously identified pathways and to identify completely novel targets and pathways.

**Table 15: The mutant strains identified during the discriminative (trained) screen that were confirmed visually. All of these mutant animals are currently being analyzed for greater genetic detail.**

| Strain Name (allele name) | Screen Method | Potential pathway |
|---|---|---|
| GT81 (a081) | Discriminative | Wnt? |
| GT82 (a082) | Discriminative | Wnt? |
| GT83 (a083) | Discriminative | Unknown |
| GT84 (a084) | Discriminative | Assembly mutants? |
| GT85 (a085) | Discriminative | Novel Class |
| GT86 (a086) | Discriminative | Wnt |
| GT87 (a088) | Discriminative | Novel Class |
| GT88 (a088) | Discriminative | Unknown |
| GT89 (a089) | Discriminative | Wnt |
| GT90 (a090) | Discriminative | Assembly mutants |

In Table 16, the animals identified during the outlier detection screen are shown. Because this screen was performed very recently, the amount of information for each of the animals is reduced relative to the discriminative screen.

**Table 16: The mutant strains identified during the outlier detection (generative) screen that were confirmed visually. All of these mutant animals are currently being analyzed for greater genetic detail.**

| Strain Name (allele name) | Screen Method | Potential pathway |
|---|---|---|
| GT98 (a098) | Outlier | Unknown |
| GT100 (a100) | Outlier | Unknown |
| GT101(a101) | Outlier | Wnt/Dshvld |
| GT102 (a102) | Outlier | Vesicle assembly? |
| GT103 (a103) | Outlier | Unknown |
| GT104 (a104) | Outlier | Wnt |
| GT105 (a105) | Outlier | Unknown |
| GT106 (a106) | Outlier | NO |
| GT107 (a107) | Outlier | Unknown |
| GT108 (a108) | Outlier | Unknown |
| GT109 (a109) | Outlier | Unknown |
| GT110 (a110) | Outlier | Unknown |
| GT112 (a112) | Outlier | Vesicle assembly? |
| GT113 (a113) | Outlier | Vesicle assembly? |
| GT114 (a114) | Outlier | Unknown |
| GT115 (a115) | Outlier | Unknown |
| GT116 (a116) | Outlier | Vesicle assembly |
| GT117 (a117) | Outlier | Differentiation |

## AUTOMATED MUTANT DISCOVERY PERFORMANCE

The difficulty of performing genetic screens based on dim or subtle fluorescent patterns is one of the major problems facing *C. elegans* neurobiology. Performing large scale forward genetic screens is the only way to create new alleles, but because it is a probabilistic process, there is uncertainty regarding the investment in time and energy. For most interesting problems some key genes can be identified by performing small forward screens or using a candidate approach to investigate a small group of genes. The Shen Lab has used candidates and small scale screens to identify genes involved in DA9 synaptogenesis, and this work resulted in papers in Cell(20), Nature(21) and Nature Neuroscience(148) within the three year period from 2007 to 2010. The exhaustion of all interesting gene candidates, however, meant that future work would have to rely heavily on genetic screens. In spite of the earlier success, and predictions that at least 15-20 genes were still unknown, it was difficult to persuade postdoctoral students to stake their careers on the results of a probabilistic forward screen (personal

communication, Kang Shen). Time consuming screens could result in merely extracting new alleles of previously implicated genes, an interesting but low-impact event, and the more genes identified as members of a specific network, the more likely this would become.

The proof of concept screens in this section only required several weeks to perform, but identified 28 mutants of interest, including three novel classes. Because of the difficulty of performing these screens manually, this could have taken a postdoctoral student over a year to complete. The success of these preliminary screens in discovering both a substantial number of novel mutants and new classes portends two significant changes to *C. elegans* neurobiology:

1) Saturation of forward screens within specific networks. Challenging screens that would have required years to saturate could be performed within a short period of time (<1 year) using this system.
2) New screening opportunities based on subtle fluorescent patterns. Screening is currently limited by the patterns changes that can be detected by the human user. Several animals identified during these screens would not have been identified during standard visual screens (personal communication, Kang Shen). Improvements in the high-content phenotyping could allow the discovery of genetic players that would have previously been overlooked.

# CHAPTER 7: CONCLUSION, THESIS CONTRIBUTIONS AND FUTURE DIRECTIONS

This dissertation focused on using recent developments in the fields of microfluidics and computer vision, and applying them to develop novel tools for biological problems. In particular, this work focused on the creation of a unified system for screening *C. elegans* to empower neurobiology research by enabling screens that would have previously been too slow and labor intensive to prove feasible. The objective was to create a system that could operate autonomously, without any human intervention, and image fluorescent reporters at the sub-cellular level, extract features to phenotype animals and identify whether the animal was a mutant of interest. Prior to this work, no one had demonstrated the ability to autonomously screen a mutagenized population of multicellular model organisms and identify novel mutants. Meeting this objective required creating a microfluidic device for animal handling, external components for controlling the system, and a computer vision framework to phenotype animals.

The microfluidic device was created by optimizing an existing device developed by Dr. Kwanghun Chung. The first generation device created by Dr. Chung allowed animals to be immobilized, imaged and sorted, but was prone to a high failure rate, inconsistent operation and a mediocre throughput. To allow for automated operation, this device was redesigned to dramatically increase the fabrication yield, as well as to increase the consistency of device operation and allowing for error handling routines. Specifically, changes to the device increased the fabrication tolerances (alignment and bonding) and nearly doubled to device yield. Partial closure valves increased the size of the smallest feature in the device, and reduced the failure rate due to dust particles in addition to simplifying the fabrication. The addition of a higher pressure fluid to fluid to flush animals out allowed animals to exit consistently, and provided a mechanism to handle device errors, such as clogging or resistance changes, without human intervention. Redesigning the cooling channel reduced the temperature gradient experienced by the animal and increased the consistency of device operation.

The external control system and components were engineered to allow computerized control of the device and to allow for consistent operation and immobilization of the animals. The previous system in place required manual, physical actuation of control lines and had no computerized implementation or ability. Specific components were created to allow actuation of pressure lines, control of solenoid valves and on-chip temperature control. This allowed error handling routines to be developed that made extended operation without human intervention possible. Not only did these components allow the automated screening, but they have enabled numerous other projects in the Lu Lab.

The computer vision and phenotyping framework was created to allow for automated image processing to identify subtle changes in fluorescent expression and thus discover novel genes affecting synapotgenesis. The computer vision system, designed using recent computer vision advances, was capable to identifying subcellular synaptic features even in an extremely noisy background, with a very low error rate. Quantitative features were then extracted from the identified features and used to create a robust phenotypical space. This framework allowed forward genetic screening to identify numerous mutants with novel synaptic expression in a completely autonomous manner. These novel mutants are in the process of being sequenced, but several novel genes have already been identified along with numerous new alleles.

In summary, this thesis makes several major contributions:

1) The creation of an optimized microfluidic device for robust handling of *C. elegans* to enable imaging and sorting at high-magnification. This device performs consistently, has a higher fabrication yield and can operate continuously without human intervention.
2) Engineering and construction of external components to allow for computerized control of device operation, immobilization and error handling routines.

3) The development of a computer vision framework to identify subtle fluorescently labeled features (synapses, cell-bodies) and separate them from surrounding tissues and autofluorescence.
4) The creation of quantitative features that allow for the creation of a quantitative phenospace incorporating more features than alternative phenotyping methodologies in neurobiology.
5) Demonstrating the utility and power of the combined system by performing the first large-scale automated forward genetic screens to identify novel genes involved in synapse formation and guidance. Found several novel genes, numerous new alleles, and additional animals are currently being sequenced.

In Table 17, some of the contributions made by this work are summarized. Of particular note is the fact that the first genetic screen of a multi-cellular model organism in a microfluidic system was done in this work. Following this, the first automated genetic screen of a multi-cellular model organism in a microfluidic system was performed. These screens identified several novel genes in addition to multiple alleles of previously discovered genes. Further sequencing is currently being performed on the additional mutants.

**Table 17: Specific contributions to the development of an autonomous, high-content screening system for *C. elegans* biology. Prior to my collaboration with Dr. Chung, there was no method of computerized device control, and thus several of these measurements were not applicable prior to my work. For comparison, however, the values achieved during the collaborative work are listed and labeled as such.**

| | Current Methods (Manual Microscopy) | Contemporaries | Thesis work |
|---|---|---|---|
| Fabrication Yield | N/A | 50% (1st Generation Device) | >90% |
| Computer control of on-chip valves | N/A | No | Yes |
| Temperature gradient across animal | N/A | 7°C(66) | 1°C |
| MTBF of a running device (requiring device change) | N/A | 1.5 hours | 7 hours |
| MTBF of a running device (requiring human intervention) | ~2 hours of imaging between breaks | 15 mins | 4.5 hours |
| Automated sorting | No | No | Yes |
| Method for identifying reporter | | Simple thresholding (57, 59, 66) | Multi-layer feature extraction and computer vision |
| Throughput of high-resolution sorting of reporter in head or tail | <100 | 60 worms/hour(66, 149) | 240 worms/hour |
| Automated screening to identify novel mutants | No | No | Yes |
| Forward genetic screening results | Yes | N/A | 2 novel genes (sequenced)<br>21 mutants w/ altered phenotypes (visual screens)<br>28 mutants w/ altered phenotypes (automated screens) |

## FUTURE DIRECTIONS

Although genetic screens of multicellular model organisms are an important part of modern biology, the every-day common methods with which scientists manipulate these organisms are still labor intensive and time-consuming. This makes many interesting projects unfeasible due to the labor and associated costs. High-throughput and high-resolution microscopy and phenotyping could eliminate a significant bottleneck in genetic analyses and enables large-scale quantitative experimentation in developmental biology, functional genomics, network biology, and other related fields. More importantly, the approach of combining automation, computer vision and microfluidics should allow for rapid complex genetic screens based on subtle phenotypes that would be otherwise difficult or impossible to detect. For example, human eyes are imprecise at detecting absolute changes in brightness; by using this approach, one can potentially screen for mutants that have altered intensity of reporters or that have slightly altered morphology. In gene expression analysis, it can also drastically improve the quality of expression pattern data due to the less subjective nature of this approach. The opportunity therefore exists, not merely to reduce the labor or cost of a project, but to perform extract phenotypical information that would be impossible by eye.

The use of microfluidic engineering methods presents excellent opportunities in making an impact in these fields for high-throughput. To best accomplish this, understanding the needs of the biology communities is crucial, as is the ability to harness the unique advantages conferred by microfluidics in terms of manipulating flow and the transport of mass, energy, and momentum. Several features developed in this project could be easily applied to other systems. It is fast, allowing processing of large numbers of animals for genomic-scale studies. These design principles could be used and adapted for *C. elegans* of different sizes and potentially other small organisms (e.g. *Drosophila melanogaster* and *Danio rerio* embryos). To truly impact the biology community further work, or commercialization, would have to be done to enable transferring these tools to end-users.

In particular I believe that the future of microfluidics for multicellular organisms will focus on standardization, increased scale of experimentation and automation. For instance, many large-scale experiments such as genetic screen (in developmental biology) or drug screens for particular disease models (e.g. Alzheimer's disease) will benefit from automation and high-throughput. The many laboratories performing related experiments and laboratories mining the large-scale data sets for genetic or genomic studies will also want to be able to compare data gathered from different experiments and from different laboratories, making standardization an absolute necessity.

The use of computer vision methods applied to cutting edge biology presents similar opportunities for making a significant impact for high-content screens. The use of computer vision methods to extract specific quantitiative features from biological images is still in its relative infancy, and promises to be a high impact area. The specific imaging constraints and computer vision approach in order to minimize error while meeting the time and computational constraints must be considered. When properly designed, these computer vision tools promise to speed discovery, and to identify features and connections that would have previously been impossible.

The application of microfluidics to enable high-throughput screening in a controlled manner and couple with computer vision to allow high-content and automation will significantly alter the way biology is currently performed. I eagerly anticipate the impacts these engineered microsystems will have in fundamental genetics and disease studies, systems biology, and pharmaceutical developments.

# APPENDIX A: PUBLICATIONS AND OTHER SCIENTIFIC ACTIVITIES

## *Journal*

**Jeffrey N. Stirman, Matthew M. Crane, Steven J. Husson, Christian Schultheis, Alexander Gottschalk, and Hang Lu, "Real-time multimodal optical control of individual neurons and muscles in freely behaving Caenorhabditis elegans", Nature Methods, 20011, 2: 153-159.**

**Matthew M. Crane, Kwanghun Chung, Jeffrey Stirman, and Hang Lu, "Microfluidics-enabled phenotyping, imaging, and screening of multicellular organisms", Lab on a Chip, 2010, 10: 1509-1517.**

**Matthew M. Crane, Kwanghun Chung, and Hang Lu, "Computer-enhanced high-throughput genetic screens of *C. elegans* in a microfluidic system", Lab on a Chip, 2009, 9 (1): 38-40.**

**Kwanghun Chung*, Matthew M. Crane*, Hang Lu, "Automated on-Chip Rapid Microscopy, Phenotyping, and Sorting of *C. elegans*", Nature Methods, 2008, 5: 637-643.**

## *Conference Talks*

**Matthew M. Crane, Jeffrey Stirman, and Hang Lu. "Autonomous synaptogenesis screening via SVM-generated quantitative phenotypical space" CSHL 2010 Meeting: Automated Imaging & High-Throughput Phenotyping.**

**Jeffrey Stirman, Matthew M. Crane, Alexander Gottschalk, and Hang Lu. Spatial and temporal optical activation of neurons in freely behaving worms" Neuroscience Topic Worm Meeting, 2010.**

**Kwanghun Chung, Jeffrey Stirman, Matthew M. Crane, and Hang Lu. "Large-scale in vivo genetic screens and laser microsurgery enabled by automated Microsystems." Association for Lab Automation, 2010.**

**Matthew M. Crane and Hang Lu. "Machine learning for spatiotemporal analysis of transcriptional regulation through analysis of GFP expression in multicellular organisms" BMES, 2009**

**Matthew M. Crane, Kwanghun Chung, and Hang Lu. "Automated on-Chip Rapid Microscopy, Phenotyping and Sorting of *C. elegans*" Association for Lab Automation, 2009.**

**Kwanghun Chung, Matthew M. Crane, and Hang Lu. "Automated and Integrated Microsystem For High-Resolution Imaging and High-Throughput Sorting of *C. elegans*" MicroTAS, 2007.**

### *Conference Posters*

**Sharon B. Sann, Matthew M. Crane, Hang Lu, Yishi Jin. "The role of RABX-5 in regulating RAB-5 endosomal compartments and synaptic vesicle formation in *C. elegans*". Neuroscience. 2010.**

**Sharon B. Sann, Matthew M. Crane, Alicia Arney, Hang Lu, Yishi Jin. "Screen for regulators of RAB-5 endosomal compartments in synapse formation". International *C. elegans* Meeting. 2009.**

**Yunli He, Amita Patil, Matthew M. Crane, and Steven L. Garverick. "Low-Power FSK Transmitter for High-Temperature Sensors using Si Tunnel Diode and SOI MOS Varactor" INSS, 2005.**

### *Patent Applications*

**"Systems and methods for high-throughput detection and sorting," Kwanghun Chung, Matthew Crane, and Hang Lu, US Patent Office, PCT Patent Application, PCT/US2008/76869, filed September 18, 2008.**

**"Real-time multi-spectral optical illumination of model organisms or cells," Jeffrey Stirman, Matthew Crane, and Hang Lu, US Patent Office, Provisional Patent, GTRC #5514, filed January 14, 2011.**

# APPENDIX B: MICROFLUIDIC FABRICATION METHODS

The creation of a master mold is the most important step of the soft-lithography process. Typically a design for the mold is created using CAD software such as AutoCad. This design is then translated and printed on a high-resolution transparency. A high-resolution transparency (20,000 dpi) typically can make channels down to 10 microns; if a higher resolution mask is needed, this can be made using a chrome-glass mask. In practice, this is rarely done because of the increased cost and that a 10 micron resolution is typically more than sufficient for most devices. The resulting mask has regions where it is clear to allow light through to the photoresist, and regions where it is black to prevent any light from being transmitted.

Once a mask has been created, the master can be fabricated in a clean-room. Because of the required resolution, even small dust particles can destroy a master, and thus fabrication should be performed in a standard class 100 fabrication facility. Using a blank silicon wafer, a positive or negative photoresist is poured onto the wafer and spun-coat to the desired thickness (ultimately corresponding to the height of the channel). The mask is then used to selectively expose parts of the photoresist, and create the ultimate three dimensional mold on top of the silicon wafer. Post exposure, the wafer is placed in a chemical developer that removes the undesired extra photoresist that surrounds the new structures.

There are two types of photoresists that can be used: positive and negative. In a negative photoresist, exposure to UV light causes the polymer to cross-link. After exposure, the areas of the wafer that weren't exposed to light wash off in the chemical developer. By contrast, the cross-linking in a positive photoresist is destroyed when exposed to light. It is these, light exposed areas that are then removed when immersed in a chemical developer. The most popular negative photoresists for soft-lithography are the SU product line from MicroChem. These resists are can be used to create extremely high aspect ratio structures, and replica molded many times without problems. The most common positive photoresist is the AZ line, also from MicroChem. In general, positive photoresists aren't as robust as negative photoresists, and need to

be handled more carefully. One significant advantage of positive photoresists over negative photoresists, however, is the ability to reflow the resist to create alternative structures. This is used to create alternative cross-sections in the channels, and allow for valves that close completely.

Once the master has been fabricated, it is placed in a dessicator and small volume of tricholoro-silane is placed in a glass vial. A vacuum is then pulled on the dessicator, which causes the silane to evaporate. Leaving the wafer under vacuum for 6-8 hours causes a thin layer of silane to be deposited on the wafer. This reduces the adhesion between the PDMS and the photoresist master and prevents the photoresist from being peeled off.

The master fabrication allows the user to then mold hundreds of devices from a single mold. The PDMS pre-polymer is mixed with the cross-linker at the appropriate ratio. The standard ratio is a weight ratio of 10:1 pre-polymer:cross-linker. Increasing the amount of cross-linker by mixing at 5:1 increases the Young's modulus and creates a significantly stiffer polymer. Conversely, reducing the mixing ratio to 15:1 or 20:1 creates a polymer with a much lower Young's modulus. All devices described in this dissertation use a mixing ratio of 10:1. The combined pre-polymer and cross-linker is stirred robustly and then placed into a vacuum chamber to allow for rapid degassing of any bubbles trapped in the mixture. Following degassing, the wafer is placed in a small container and the polymer is slowly poured over the master. Depending on the smallest feature size, air bubbles may be trapped in the master if the polymer is poured too quickly onto the wafer, and thus require additional degassing in a vacuum chamber. To fully cure the PDMS, the container is then placed in a 70C oven for two hours, or stored for 36-48 hours at room temperature.

Multi-layer devices are important as they allow one of the primary advantages soft-lithography (a simple valving mechanism), as well as allowing the creation of complex chemical/gas/thermal gradients in devices using diffusion through PDMS. Multilayer devices are typically fabricated using a thick layer made of bulk PDMS poured over a master mold, and a thin layer formed by spin-coating PDMS over a separate

mold. To form the thick layer, PDMS is poured over the wafer to create a PDMS layer one the order of 1 cm tall, and is partially cured (~25 mins at 70C). The thick layer is then removed from the mold. The thin layer is formed by spin-coating a small amount of PDMS on top of the mold. Varying the spin speed allows creation of a PDMS film with a specific height. This wafer is allowed to reflow for 30 mins to ensure that the film height is uniform, and placed on a hot plate set at 65C for 10 mins. The thick layer devices are then aligned to the thin layer underneath a stereomicroscope. Because both layers have only been partially cured, the polymer chains will bond across the interface of the two layers and create a strong bond. The combined layers should then be allowed to finish curing overnight.

In order to create sealed channels, the molded device, must then be bonded to a glass coverslip or slide. This is typically done by immersing the entire PDMS device in oxygen plasma for a short period of time. This process activates the oxygen radicals on the surface of the PDMS, which, when placed on a piece of glass, covalently bond with the glass. The bond created by this is irreversible and extremely strong, and forms in seconds. Sealing a device in this manner takes only a few minutes and is capable of withstanding pressures greater than 70psi. This plasma-bonding process can also be used to bond PDMS to itself, or to other materials such as quartz, silicon nitride and polystyrene, but in practice is most commonly used to bond to glass slides or coverslips – depending on the choice of microscope objective.

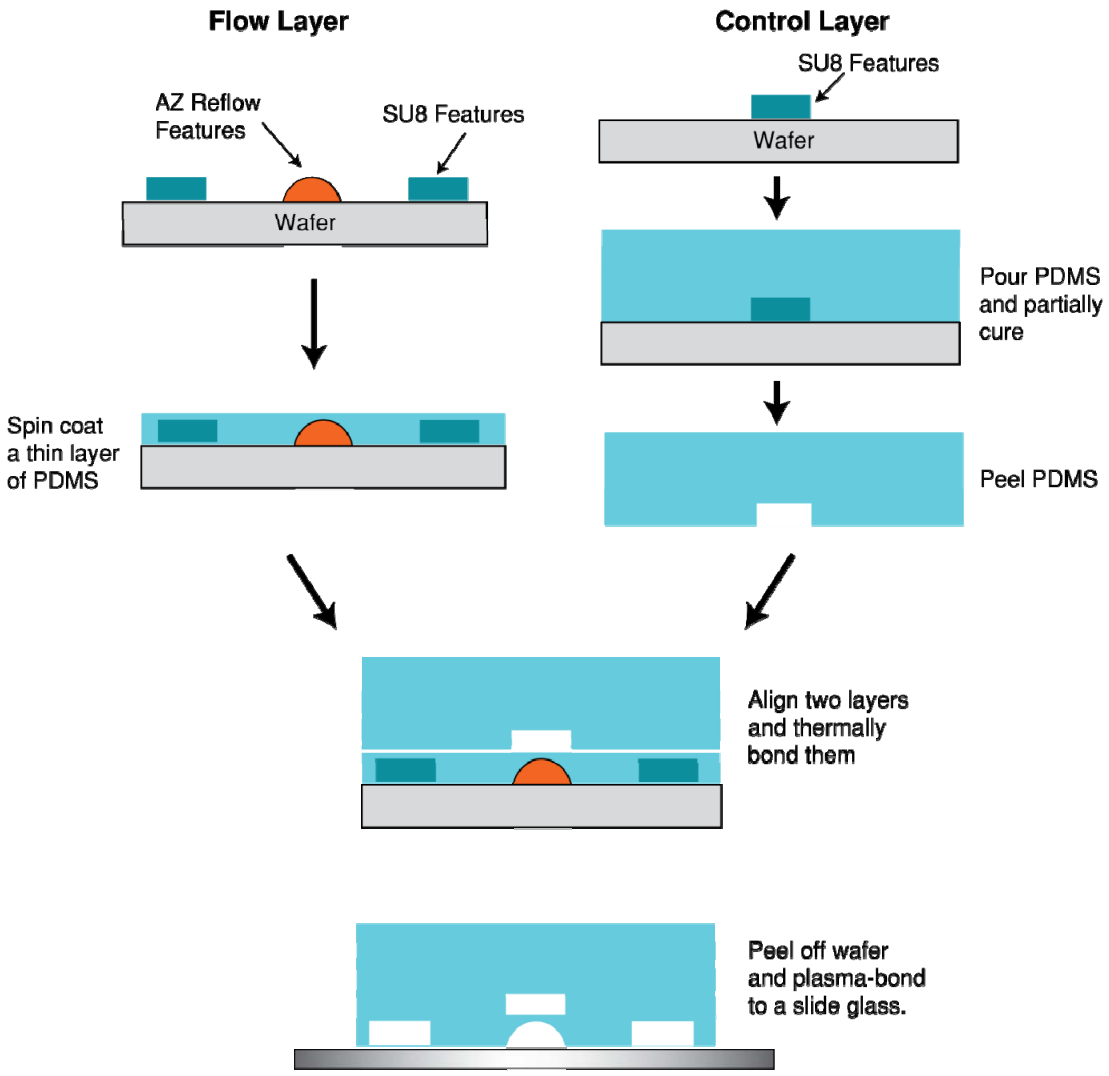



**Figure 62: The two-layer fabrication process. This diagram shows the steps required to create a two-layer device. A small amount of PDMS is spun coat onto the thin layer, which is partially cured. A thick (~3mm) layer of PDMS is partially cured on the control layer. The thick layer devices are removed from the wafer, cut, and then aligned to the thin-layer. They are then bonded to slide glass.**

# APPENDIX C: TROUBLESHOOTING METHODS DEVELOPED

Silane treatment: If the silane treatment of the master doesn't appear to be working, in that the PDMS is hard to remove from the wafer and/or removes the photoresist, this is likely a result of ineffective silane treatment. This is likely a result of either silane that has been left open and exposed to the ambient for an extended period of time, or an insufficiently strong vacuum. Test the vacuum desiccator to ensure that it is pulling a strong vacuum, and if it is – purchase new silane.

Bonding issue: If the bonds between layers in the microfluidic device don't appear sufficiently strong, there could be several culprits. If the bond between the layers that were partially cured and then bonded appears week, the probably culprit is over curing one of the layers. Repeat the fabrication process but bake the thin, spun-coated layer for a few minutes less. If the bond between layers that were plasma activated is weak, there could be several causes. First turn on the plamsa asher and run for ~20 minutes. This will clean the machine and remove any organic substances that would be present and interfering with the oxygen plasma. Then treat pieces of PDMS for different amounts of time and attempt to bond them together. Experiment with both less time, and more time until you find the proper settings for your machine.

Leakage: Leakage typically results from one of two sources: bonding issues, or a poor seal between the device and connected tubing/pins. If the cause seems to be bonding, see the bonding troubleshooting entry. If the problem appears to be a result of a bad seal at between the connected pins and the device, it could be as a result of either incorrectly sized pins or a tearing around the holes in the device. The silicone material of the device is used to form a strong seal around pins and tubing connected to it. When the holes in the device are punched, however, it is possible to do this carelessly and create small tears around the hole which prevents a good seal from forming. If this is the case, make sure to take care when punching holes, and possibly sharpen the pins used for punching holes in the device.

# APPENDIX D: WORM PROTOCOLS

Animals were cultured according to established methods(150). Mutagenesis was performed on age-synchronized L4 animals using EMS according to standard protocols.(151) F2 eggs were obtained by bleaching F1 adults using a solution containing about 1% NaOCl and 0.1 M NaOH, washed in M9 buffer, and cultured on Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50 until L4 stage.

For the preliminary screens using the simplified microfluidic device discussed in Chapter 2, the following procedure was used. Animals were washed and suspended in M9 solution containing 0.02 wt% Bovine Serum Albumin (BSA) for each experiment. Animals were screened under a compound microscope at 20X based on differences in the reporter expression pattern or intensity; potential animals of interest were sorted into the mutant outlet and were collected directly from tubing connected to the mutant outlet with M9 solution containing 0.02 wt% BSA. Animals were subsequently transferred to individual plates for culture and further examination. The first automated sorting(66) used an identical procedure, but using the objectives disclosed in the text.

For the later experiments including the automated screening discussed in Chapter 6, the procedure was modified to eliminate the BSA. Animals were washed and suspended in M9 solution containing 0.01 volume% Triton X100 for each experiment. This prevented the animals from sticking to the tubing during injection. Animals were screened under a compound microscope using a 40X oil objective. Sorting decisions were made based on differences in the reporter expression pattern or intensity; potential animals of interest were sorted into the mutant outlet and were collected directly from tubing connected to the mutant outlet with M9 solution containing 0.02 wt% BSA. Animals were subsequently transferred to individual plates for culture and further examination.

## REFERENCES


1. Crane MM, Chung K, & Lu H (2009) Computer-enhanced high-throughput genetic screens of C. elegans in a microfluidic system. *Lab on a Chip* 9(1):38-40.

2. Kaletta T & Hengartner MO (2006) Finding function in novel targets: C-elegans as a model organism. *Nature Reviews Drug Discovery* 5(5):387-398.

3. Lai CH, Chou CY, Ch'ang LY, Liu CS, & Lin WC (2000) Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics. (Translated from English) *Genome Research* 10(5):703-713 (in English).

4. Kuwabara PE & O'Neil N (2001) The use of functional genomics in C. elegans for studying human development and disease. *Journal of Inherited Metabolic Disease* 24(2):127-138.

5. Harris TW*, et al.* (2004) WormBase: a multi-species resource for nematode biology and genomics. (Translated from English) *Nucleic Acids Research* 32:D411-D417 (in English).

6. Sulston JE & Horvitz HR (1977) Post-Embryonic Cell Lineages of Nematode, Caenorhabditis-Elegans. *Developmental Biology* 56(1):110-156.

7. Fire A*, et al.* (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391(6669):806-811.

8. Chalfie M, Tu Y, Euskirchen G, Ward WW, & Prasher DC (1994) Green fluorescent protein as a marker for gene expression. pp 802-805.

9. Anonymous (1998) Genome sequence of the nematode C-elegans: A platform for investigating biology. *Science* 282(5396):2012-2018.

10. Ankeny RA (2001) The natural history of Caenorhabditis elegans research. *Nat Rev Genet* 2(6):474-479.

11. Jorgensen EM & Mango SE (2002) The art and design of genetic screens: Caenorhabditis elegans. *Nat Rev Genet* 3(5):356-369.

12. White JG, Southgate E, Thomson JN, & Brenner S (1986) The structure of the nervous system of the nematode Caenorhabditis elegans. *Phil. Trans. R. Soc. Lond. B* 314:1-340.

13. Nonet ML (1999) Visualization of synaptic specializations in live C-elegans with synaptic vesicle protein-GFP fusions. *Journal of Neuroscience Methods* 89(1):33-40.

14. Jin Y (Synaptogenesis. in *WormBook*, ed The-C-elegans-Research-Community.

15. Margeta MA, Shen K, & Grill B (2008) Building a synapse: lessons on synaptic specificity and presynaptic assembly from the nematode C-elegans. *Current Opinion in Neurobiology* 18(1):69-76.

16. Zhen M & Jin Y (1999) The liprin protein SYD-2 regulates the differentiation of presynaptic termini in C. elegans. *Nature* 401(6751):371-375.

17. Hallam SJ & Jin Y (1998) lin-14 regulates the timing of synaptic remodelling in Caenorhabditis elegans. *Nature* 395(6697):78-82.

18. Ding M, Chao D, Wang G, & Shen K (2007) Spatial Regulation of an E3 Ubiquitin Ligase Directs Selective Synapse Elimination. in *Science*, pp 947-951.

19. Maro GS, Klassen MP, & Shen K (2009) A β-Catenin-Dependent Wnt Pathway Mediates Anteroposterior Axon Guidance in <italic>C. elegans</italic> Motor Neurons. *PLoS ONE* 4(3):e4690.

20. Klassen MP & Shen K (2007) Wnt Signaling Positions Neuromuscular Connectivity by Inhibiting Synapse Formation in C. elegans. *Cell* 130(4):704-716.

21. Poon VY, Klassen MP, & Shen K (2008) UNC-6/netrin and its receptor UNC-5 locally exclude presynaptic components from dendrites. *Nature* 455(7213):669-U668.

22. Burbea M, Dreier L, Dittman JS, Grunwald ME, & Kaplan JM (2002) Ubiquitin and AP180 Regulate the Abundance of GLR-1 Glutamate Receptors at Postsynaptic Elements in C. elegans. *Neuron* 35(1):107-120.

23. Sieburth D, Madison JM, & Kaplan JM (2007) PKC-1 regulates secretion of neuropeptides. *Nature Neuroscience* 10(1):49-57.

24. Vashlishan AB*, et al.* (2008) An RNAi screen identifies genes that regulate GABA synapses. *Neuron* 58(3):346-361.

25. Ch'ng Q, Sieburth D, & Kaplan JM (2008) Profiling Synaptic Proteins Identifies Regulators of Insulin Secretion and Lifespan. *Plos Genetics* 4(11).

26. Sieburth D*, et al.* (2005) Systematic analysis of genes required for synapse structure and function. *Nature* 436(7050):510-517.

27. Bakal C, Aach J, Church G, & Perrimon N (2007) Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* 316(5832):1753-1756.

28. Peng HC (2008) Bioimage informatics: a new area of engineering biology. *Bioinformatics* 24(17):1827-1836.

29. Jones TR*, et al.* (2009) Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proceedings of the National Academy of Sciences of the United States of America* 106(6):1826-1831.

30. Conrad C & Gerlich DW (2010) Automated microscopy for high-content RNAi screening. *Journal of Cell Biology* 188(4):453-461.

31. Cohen AR, Gomes F, Roysam B, & Cayouette M (2010) Computational prediction of neural progenitor cell fates. *Nature Methods* 7(3):213-U275.

32. Cohen AA*, et al.* (2008) Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science* 322(5907):1511-1516.

33. El-Ali J, Sorger PK, & Jensen KF (2006) Cells on chips. *Nature* 442(7101):403-411.

34. Whitesides GM (2006) The origins and the future of microfluidics. *Nature* 442(7101):368-373.

35. Xia Y & Whitesides G (1998) Soft lithography. *Ann. Rev. Mater. Sci.* 28:153-184.

36. Duffy DC, McDonald JC, Schueller OJA, & Whitesides GM (1998) Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Analytical Chemistry* 70(23):4974-4984.

37. Jackman RJ, Duffy DC, Cherniavskaya O, & Whitesides GM (1999) Using elastomeric membranes as dry resists and for dry lift-off. (Translated from English) *Langmuir* 15(8):2973-2984 (in English).

38. Unger MA, Chou HP, Thorsen T, Scherer A, & Quake SR (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* 288(5463):113-116.

39. Thorsen T, Maerkl SJ, & Quake SR (2002) Microfluidic large-scale integration. *Science* 298(5593):580-584.

40. Ottesen EA, Hong JW, Quake SR, & Leadbetter JR (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* 314(5804):1464-1467.

41. Park ES, Brown AC, DiFeo MA, Barker TH, & Lu H (2010) Continuously perfused, non-cross-contaminating microfluidic chamber array for studying cellular responses to orthogonal combinations of matrix and soluble signals. *Lab on a Chip* 10(5):571-580.

42. Hirsch AM, Rivet CA, Zhang B, Kemp ML, & Lu H (2009) Parallel multi-time point cell stimulation and lysis on-chip for studying early signaling events in T cell activation. *Lab on a Chip* 9(4):536-544.

43. Shoji S (1998) Fluids for sensor systems. *Microsystem Technology in Chemistry and Life Science,* Topics in Current Chemistry, (Springer-Verlag Berlin, Berlin 33), Vol 194, pp 163-188.

44. Goll C*, et al.* (1996) Microvalves with bistable buckled polymer diaphragms. (Translated from English) *Journal of Micromechanics and Microengineering* 6(1):77-79 (in English).

45. Yang X, Grosjean C, Tai YC, & Ho CM (1998) A MEMS thermopneumatic silicone rubber membrane valve. (Translated from English) *Sens. Actuator A-Phys.* 64(1):101-108 (in English).

46. Chronis N, Zimmer M, & Bargmann CI (2007) Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans. *Nature Methods* 4:727-731.

47. Zhang Y, Lu H, & Bargmann CI (2005) Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. *Nature* 438(7065):179-184.

48. Gray JM*, et al.* (2004) Oxygen sensation and social feeding mediated by a C-elegans guanylate cyclase homologue. *Nature* 430(6997):317-322.

49. Lockery SR*, et al.* (2008) Artificial dirt: Microfluidic substrates for nematode neurobiology and behavior. (Translated from English) *J. Neurophysiol.* 99(6):3136-3143 (in English).

50. Park S*, et al.* (2008) Enhanced Caenorhabditis elegans locomotion in a structured microfluidic environment. (Translated from English) *PLoS One* 3(6):e2550 (in English).

51. Gray JM*, et al.* (2004) Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. (Translated from English) *Nature* 430(6997):317-322 (in English).

52. Zimmer M*, et al.* (2009) Neurons Detect Increases and Decreases in Oxygen Levels Using Distinct Guanylate Cyclases. (Translated from English) *Neuron* 61(6):865-879 (in English).

53. Fakhoury JR, Sisson JC, & Zhang XJ (2009) Microsystems for controlled genetic perturbation of live Drosophila embryos: RNA interference, development robustness and drug screening. (Translated from English) *Microfluidics and Nanofluidics* 6(3):299-313 (in English).

54. Lucchetta EM, Lee JH, Fu LA, Patel NH, & Ismagilov RF (2005) Dynamics of Drosophila embryonic patterning network perturbed in space and time using microfluidics. *Nature* 434(7037):1134-1138.

55. Lucchetta EM, Munson MS, & Ismagilov RF (2006) Characterization of the local temperature in space and time around a developing Drosophila embryo in a microfluidic device. (Translated from English) *Lab Chip* 6(2):185-190 (in English).

56. Dagani GT*, et al.* (2007) Microfluidic self-assembly of live Drosophila embryos for versatile high-throughput analysis of embryonic morphogenesis. (Translated from English) *Biomed. Microdevices* 9(5):681-694 (in English).

57. Chung K & Lu H (2009) Automated high-throughput cell microsurgery on-chip. (Translated from English) *Lab Chip* 9(19):2764-2766 (in English).

58. Chung KH, Crane MM, & Lu H (2008) Automated on-chip rapid microscopy, phenotyping and sorting of C. elegans. (Translated from English) *Nature Methods* 5(7):637-643 (in English).

59. Rohde CB, Zeng F, Gonzalez-Rubio R, Angel M, & Yanik MF (2007) Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution. *PNAS* 104(35):13891-13895.

60. Guo SX*, et al.* (2008) Femtosecond laser nanoaxotomy lab-on-achip for in vivo nerve regeneration studies. (Translated from English) *Nature Methods* 5(6):531-533 (in English).

61. Zeng F, Rohde CB, & Yanik MF (2008) Sub-cellular precision on-chip small-animal immobilization, multi-photon imaging and femtosecond-laser manipulation. (Translated from English) *Lab Chip* 8(5):653-656 (in English).

62. Chokshi TV, Ben-Yakar A, & Chronis N (2009) CO2 and compressive immobilization of C. elegans on-chip. (Translated from English) *Lab Chip* 9(1):151-157 (in English).

63. Quake SR & Scherer A (2000) From micro- to nanofabrication with soft materials. *Science* 290(5496):1536-1540.

64. Hulme SE, Shevkoplyas SS, Apfeld J, Fontana W, & Whitesides GM (2007) A microfabricated array of clamps for immobilizing and imaging C. elegans. *Lab on a Chip* 7(11):1515-1523.

65. Allen PB*, et al.* (2008) Single-synapse ablation and long-term imaging in live C. elegans. (Translated from English) *J. Neurosci. Methods* 173(1):20-26 (in English).

66. Chung K, Crane MM, & Lu H (2008) Automated on-chip rapid microscopy, phenotyping and sorting of C. elegans. *Nat Meth* 5(7):637-643.

67. Swets DL & Weng JJ (1996) Using discriminant eigenfeatures for image retrieval. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 18(8):831-836.

68. Turk M & Pentland A (1991) EIGENFACES FOR RECOGNITION. *Journal of Cognitive Neuroscience* 3(1):71-86.

69. Belhumeur PN, Hespanha JP, & Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19(7):711-720.

70. Zhang J, Yan Y, & Lades M (1997) Face recognition: Eigenface, elastic matching, and neural nets. *Proceedings of the IEEE* 85(9):1423-1435.

71. Moghaddam B, Jebara T, & Pentland A (2000) Bayesian face recognition. *Pattern Recognition* 33(11):1771-1782.

72. He XF, Yan SC, Hu YX, Niyogi P, & Zhang HJ (2005) Face recognition using Laplacianfaces. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 27(3):328-340.

73. Brunelli R & Poggio T (1993) FACE RECOGNITION - FEATURES VERSUS TEMPLATES. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 15(10):1042-1052.

74. Sung KK & Poggio T (1998) Example-based learning for view-based human face detection. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 20(1):39-51.

75. Papageorgiou CP, Oren M, & Poggio T (1998) A general framework for object detection. *Computer Vision, 1998. Sixth International Conference on*, pp 555-562.

76. Mohan A, Papageorgiou C, & Poggio T (2001) Example-based object detection in images by components. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23(4):349-361.

77. Heisele B, Ho P, & Poggio T (2001) Face recognition with support vector machines: global versus component-based approach. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, pp 688-694 vol.682.

78. Viola P & Jones MJ (2004) Robust real-time face detection. *International Journal of Computer Vision* 57(2):137-154.

79. Boland MV, Markey MK, & Murphy RF (1998) Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* 33(3):366-375.

80. Huang K & Murphy RF (2004) Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *Bmc Bioinformatics* 5.

81. Bao ZR*, et al.* (2006) Automated cell lineage tracing in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* 103(8):2707-2712.

82. Ranzato M*, et al.* (2007) Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters* 28(1):31-39.

83. Dankert H, Wang LM, Hoopfer ED, Anderson DJ, & Perona P (2009) Automated monitoring and analysis of social behavior in Drosophila. *Nature Methods* 6(4):297-303.

84. Branson K, Robie AA, Bender J, Perona P, & Dickinson MH (2009) High-throughput ethomics in large groups of Drosophila. *Nature Methods* 6(6):451-U477.

85. Mukherjee DP, Ray N, & Acton ST (2004) Level set analysis for leukocyte detection and tracking. *Ieee Transactions on Image Processing* 13(4):562-572.

86. Debeir O, Van Ham P, Kiss R, & Decaestecker C (2005) Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *Ieee Transactions on Medical Imaging* 24(6):697-711.

87. Burges CJC (1998) A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery* 2(2):121-167.

88. Aizerman A, Braverman EM, & Rozoner LI (1964) Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control* 25:821-837.

89. Boser BE, Guyon IM, & Vapnik VN (1992) A training algorithm for optimal margin classifiers. in *Proceedings of the fifth annual workshop on Computational learning theory* (ACM, Pittsburgh, Pennsylvania, United States), pp 144-152.

90. Chang CC & Lin CJ (LIBSVM: a library for support vector machines, 2001, software available at http://www.csie.ntu.edu.tw/ ~cjlin/ libsvm.

91. Fan R-E, Chang K-W, Hsieh C-J, Wang X, & Lin C-J (2008) LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* (9).

92. Meir R & Ratsch G (2002) An introduction to boosting and leveraging. *Advanced Lectures on Machine Learning* 2600:118-183.

93. Tieu K & Viola P (2004) Boosting image retrieval. *International Journal of Computer Vision* 56(1-2):17-36.

94. Fasel I, Fortenberry B, & Movellan J (2005) A generative framework for real time object detection and classification. *Computer Vision and Image Understanding* 98(1):182-210.

95. Opelt A, Pinz A, Fussenegger M, & Auer P (2006) Generic object recognition with boosting. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 28(3):416-431.

96. Friedman J, Hastie T, & Tibshirani R (2000) Additive logistic regression: A statistical view of boosting. *Annals of Statistics* 28(2):337-374.

97. Freund Y & Schapire R (1996) Experiments with a new boosting algorithm. in *In: Machine Learning: Proceedings of the 13th International Conference, Morgan Kauffman*, pp 148--156.

98. Freund Y & Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci* (55):119--139.

99. Grove AJ & Schuurmans D (1998) Boosting in the limit: Maximizing the margin of learned ensembles. in *Proceedings of the Fifteenth National ConferenceonArtificial Intelligence*, pp 692--699.

100. Breiman L (1996) Bagging predictors. in *Machine Learning*, pp 123--140.

101. Dalal N, Triggs B, & Schmid C (2006) Human detection using oriented histograms of flow and appearance. *Computer Vision - Eccv 2006, Pt 2, Proceedings* 3952:428-441.

102. Lowe DG (1999) Object Recognition from Local Scale-Invariant Features. in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2* (IEEE Computer Society), p 1150.

103. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2):91-110.

104. Sigala R, Serre T, Poggio T, & Giese M (2005) Learning features of intermediate complexity for the recognition of biological motion. *Artificial Neural Networks: Biological Inspirations - Icann 2005, Pt 1, Proceedings* 3696:241-246.

105. Bosch A, Zisserman A, & Munoz X (2006) Scene classification via pLSA. *Computer Vision - Eccv 2006, Pt 4, Proceedings* 3954:517-530.

106. Quelhas P, Monay F, Odobez JM, Gatica-Perez D, & Tuytelaars T (2007) A thousand words in a scene. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 29(9):1575-1589.

107. Bosch A, Zisserman A, & Munoz X (2008) Scene classification using a hybrid generative/discriminative approach. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 30(4):712-727.

108. Angeli A, Filliat D, Doncieux S, & Meyer JA (2008) Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words. *Ieee Transactions on Robotics* 24(5):1027-1037.

109. Wang J*, et al.* (2008) Cellular phenotype recognition for high-content RNA interference genome-wide screening. (Translated from English) *J. Biomol. Screen* 13(1):29-39 (in English).

110. Singh DK*, et al.* (2010) Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Mol Syst Biol* 6.

111. Geng W, Cosman P, Palm M, & Schafer WR (2005) Caenorhabditis elegans egg-laying detection and behavior study using image analysis. *Eurasip Journal on Applied Signal Processing* 2005(14):2229-2240.

112. Geng W, Cosman P, Berry CC, Feng ZY, & Schafer WR (2004) Automatic tracking, feature extraction and classification of C-elegans phenotypes. *Ieee Transactions on Biomedical Engineering* 51(10):1811-1820.

113. Geng W, Cosman P, Baek JH, Berry CC, & Schafer WR (2003) Quantitative Classification and Natural Clustering of Caenorhabditis elegans Behavioral Phenotypes. *Genetics* 165(3):1117-1126.

114. Murray J*, et al.* (2008) Automated analysis of embryonic gene expression with cellular resolution in C. elegans. *Nature Methods* 5(8):703 - 709.

115. Boyle T, Bao Z, Murray J, Araya C, & Waterston R (2006) AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis. *Bmc Bioinformatics* 7(1):275.

116. Bao Z*, et al.* (2006) Automated cell lineage tracing in Caenorhabditis elegans. *PNAS* 103(8):2707 - 2712.

117. Murray J, Bao Z, Boyle T, & Waterston R (2006) The lineaging of fluorescently-labeled Caenorhabditis elegans embryos with StarryNite and AceTree. *Nature Protocols* 1(3):1468 - 1476.

118. Bargmann CI & Avery L (1995) Laser killing of cells in Caenorhabditis elegans. *Methods Cell Biol* 48:225-250.

119. Rohde C, Gilleland C, Samara C, Zeng F, & Yanik MF (2008) High-throughput in vivo genetic and drug screening using femtosecond laser nano-surgery, and microfluidics. (Translated from English) *Conf Proc IEEE Eng Med Biol Soc* 2008:2642 (in English).

120. Cornell E*, et al.* (2008) Automating fruit fly Drosophila embryo injection for high throughput transgenic studies. *Review of Scientific Instruments* 79(1):013705.

121. Lu Z, Chen PCY, Nam J, Ge RW, & Lin W (2007) A micromanipulation system with dynamic force-feedback for automatic batch microinjection. (Translated from English) *J. Micromech. Microeng.* 17(2):314-321 (in English).

122. Wang W, Liu X, Gelinas D, Ciruna B, & Sun Y (2007) A Fully Automated Robotic System for Microinjection of Zebrafish Embryos. *PLoS ONE* 2(9):e862.

123. Zappe S, Fish M, Scott MP, & Solgaard O (2006) Automated MEMS-based Drosophila embryo injection system for high-throughput RNAi screens. (Translated from English) *Lab Chip* 6(8):1012-1019 (in English).

124. Zhang XJ, Scott MP, Quate CF, & Solgaard O (2006) Microoptical characterization of piezoelectric vibratory microinjections in Drosophila embryos for genome-wide RNAi screen. (Translated from English) *J. Microelectromech. Syst.* 15(2):277-286 (in English).

125. Sun Y, Duthaler S, & Nelson BJ (2004) Autofocusing in computer microscopy: Selecting the optimal focus algorithm. *Microscopy Research and Technique* 65(3):139-149.

126. Price JH & Gough DA (1994) Comparison of Phase-Contrast and Fluorescence Digital Autofocus for Scanning Microscopy. *Cytometry* 16(4):283-297.

127. Grill B*, et al.* (2007) C. elegans RPM-1 Regulates Axon Termination and Synaptogenesis through the Rab GEF GLO-4 and the Rab GTPase GLO-1. *Neuron* 55(4):587-601.

128. Brown HM, Van Epps HA, Goncharov A, Grant BD, & Jin YS (2009) The JIP3 Scaffold Protein UNC-16 Regulates RAB-5 Dependent Membrane Trafficking at C. elegans Synapses. *Developmental Neurobiology* 69(2-3):174-190.

129. Lee JN, Park C, & Whitesides GM (2003) Solvent Compatibility of Poly(dimethylsiloxane)-Based Microfluidic Devices. *Analytical Chemistry* 75(23):6544-6554.

130. Sonnichsen B*, et al.* (2005) Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* 434(7032):462-469.

131. Sieburth D*, et al.* (2005) Systematic analysis of genes required for synapse structure and function. *Nature* 436:510-517.

132. Hunt-Newbury R*, et al.* (2007) High-Throughput In Vivo Analysis of Gene Expression in Caenorhabditis elegans. *PLoS Biol* 5(9):e237.

133. Herman RK (1989) Mosaic analysis in the nematode Caenorhabditis elegans. *J Neurogenet* 5(1):1-24.

134. Brenner S (1974) The genetics of Caenorhabditis elegans. *Genetics* 77(1):71-94.

135. Chang AJ, Chronis N, Karow DS, Marletta MA, & Bargmann CI (2006) A Distributed Chemosensory Circuit for Oxygen Preference in C. elegans. *PLoS Biol* 4(9):e274.

136. Troemel ER, Sagasti A, & Bargmann CI (1999) Lateral signaling mediated by axon contact and calcium entry regulates asymmetric odorant receptor expression in C. elegans. *Cell* 99(4):387-398.

137. Dittman JS & Kaplan JM (2006) Factors regulating the abundance and localization of synaptobrevin in the plasma membrane. *Proceedings of the National Academy of Sciences* 103(30):11399.

138. Wang M, Zhou XB, King RW, & Wong STC (2007) Context based mixture model for cell phase identification in automated fluorescence microscopy. (Translated from English) *Bmc Bioinformatics* 8:12 (in English).

139. Davis JW & Keck MA (2005) A Two-Stage Template Approach to Person Detection in Thermal Imagery. *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*, pp 364-369.

140. Nayar SK, Nene SA, & Murase H (1996) Real-time 100 object recognition system. *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, pp 2321-2325 vol.2323.

141. Bileschi S (2006) StreetScenes: Towards scene understanding in still images. (MIT EECS).

142. Plath N, Toussaint M, & Nakajima S (Multi-class image segmentation using Conditional Random Fields and Global Classification.

143. Lafferty J & McCallum A (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data. in *Proc. ICML*.

144. Wallach HM (2004) Conditional random fields: An introduction.

145. Quattoni A, Collins M, & Darrell T (2005) Conditional random fields for object recognition. in *In Advances in Neural Information Processing Systems 17*.

146. He X, Zemel RS, & Ray D (2006) Learning and incorporating top-down cues in image segmentation. in *In Proceedings of the 9th European Conference on Computer Vision*.

147. Margeta MA, Wang GJ, & Shen K (2009) Clathrin adaptor AP-1 complex excludes multiple postsynaptic receptors from axons in C. elegans. *Proceedings of the National Academy of Sciences of the United States of America* 106(5):1632-1637.

148. Teichmann HM & Shen K (2010) UNC-6 and UNC-40 promote dendritic growth through PAR-4 in Caenorhabditis elegans neurons. *Nat Neurosci* advance online publication.

149. Guo SX*, et al.* (2008) Femtosecond laser nanoaxotomy lab-on-a-chip for in vivo nerve regeneration studies. *Nat Meth* 5(6):531-533.

150. Brenner S (1974) Genetics of Caenorhabditis-Elegans. (Translated from English) *Genetics* 77(1):71-94 (in English).

151. Wood WB (1988) *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory Press, New York) p 667.